Energy efficient deployment and operation strategies for wireless networking

Submitted in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

in

Computer Science

at the

UNIVERSITY OF ALBERTA

Abstract

Computing and other Information and Communication Technology (ICT) resources lend themselves to reducing power consumption in an industrialised society. For example, holding a virtual teleconference instead of delegates travelling to a particular location for a meeting conveys a substantial energy saving. However, currently ICT power consumption is comparable with that of global air travel and continues to rise. In attempting to mitigate the effects of climate change society as a whole must strive to reduce power consumption.

Over 20% of the ICT power consumed occurs with wireless telecommunications. In this report the technology for wireless networking is reviewed together with measures designed to improve communication efficiency, usability and network transparency whilst simultaneously utilising lower power.

To supplement the literature review, a number of simulation experiments have been identified to supplement this work utilising the tools provided within the MATLAB simulation environment. Due to unexpected results encountered during initial trials of the software simulation framework these simulation experiments have not currently taken place until the problem is resolved.

Declaration

I hereby declare that I am the sole author of this work. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Andrew Patrick Kemp

....

Signature

....

Date

Acknowledgements

I would like to thank my family for their love, support, encouragement and understanding that have brought me to this point in my life. I feel especially thankful to be so fortunate as to be blessed with such a loving wife and daughter. I continue to love you both a little more than I did yesterday but a little less than I shall tomorrow. How you haven't left home whilst I have struggled to navigate and process the material for this project is something I shall never understand and something I shall forever be grateful for. Despite my best efforts I know I've been particularly unbearable to live with whilst I've been writing. To be able to look in your faces each morning and remember the past helps me to strive forwards through another day.

I would also like to thank my past supervisors, lecturers, colleagues and students for sharing their knowledge, insight and life experiences with me so freely. I look back on our discussions and remain bewildered at the the understanding I continue to gain from them.

Since coming to Canada, I have been blessed with new friends who give so freely of themselves and have provided much welcome diversion on the rare occasions I have been able to spare the time to share their company.

For all this and so much more, I am grateful to you all. Any successes I may make continue to be a testament to my knowing you. The failures and mistakes... well, they continue to be mine and mine alone...

Coffee continues to be an essential yet understated contributor to my early mornings and late nights. Thank you. Also instrumental to this work was LaTeX for typesetting. I confess I hated TeX when I first used it over 30 years ago. RTFM... (Which one? Which part?) ... Need to talk to a human? "... You need a different version/compiler/script/operating system..." Things haven't improved: I still do. I continue to pronounce the 'x' in protest - **thank you** Leslie Lamport for giving me permission to do so.

I would like to thank Chintha Tellambura for acting as my project supervisor.

I thank God for allowing me sufficient good health to complete this work. I wish however that you had provided me with the foresight to be able to pick an environment that would have allowed me to have produced work which reflected

at least a small proportion of the effort I have invested. I also thank God it's over.

Finally, dear reader, I thank you for your perseverence in reading my work thus far. You have already given me more attention than everyone else supposedly associated with this project.

Abbreviations

ACI	Adjacent Channel Interference
ACK	ACKnowledgement
AI	Artificial Intelligence
AP	Access Point
ARQ	Automatic Repeat Request
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BAM	Binary Amplitude Modulation
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
BSS	Basic Service Set
CCI	Co-Channel Interference
CDMA	Code Division Multiple Access
COW	Copy On Write
CP	Continuous Phase
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CSMA	Carrier Sense Multiple Access
CTS	Clear To Send
FHSS	Direct Sequence Spread Spectrum
ESS	Extended Service Set
FDMA	Frequency Division Multiple Access
FHSS	Frequency Hopping Spread Spectrum
FSK	Frequency Shift Keying
FSPL	Free Space Path Loss
GDP	Gross Domestic Product
GMSK	Gaussian Minimum Shift Keying
ICI	Inter Channel Interference
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronic Engineers

IETF	Internet Engineering Task Force
IFS	InterFrame Space
IMT	International Mobile Telecommunications
ISI	Inter Symbol Interference
ISM	Industrial Scientific and Medical
ISO	International Organization for Standardization
ITU	International Telecommunications Union
LAN	Local Area Network
LLC	Logical Link Control
MAC	Medium Access Control
MCS	Modulation Coding Scheme
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MS	Mobile Station
MSK	Minimum Shift Keying
OMG	Object Management Group
OO	Object Oriented
OQPSK	Offset Quadrature Phase Shift Keying
OSI	Open Standards Interconnect
PID	Proportional-Integral-Derivative
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RADAR	RAdio Detection And Ranging
RADIUS	Remote Authentication Dial In User Service
RF	Radio Frequency
RFID	Radio Frequency IDentification
RMS	Root Mean Square
RoI	Return on Investment
RTS	Ready To Send
SDMA	Space Division Multiple Access
SE	Spectral Efficiency
SEAD	Super-efficient Equipment and Appliance Deployment
SIMO	Single Input Multiple Output
SINR	Signal to Interference plus Noise Ratio

SISO	Single Input Single Output
SNR	Signal to Noise Ratio
STA	STAtion
TDD	Test Driven Development
TDMA	Time Division Multiple Access
UML	Unified Modelling Language
UN	United Nations
WAN	Wide Area Network
WPAN	Wireless Personal Area Network
WDM	Wavelength Division Multiplexing
WM	Wireless Medium

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"I am cast upon a horrible desolate Island, void of all hope of Recovery. But I am alive, and not drown'd as all my Ship'd Company was."*
>
> – Daniel Defoe, *Robinson Crusoe*

The effects of two hundred years' industrialisation upon the global environment have been clearly and carefully documented by many workers to form an indisputable body of evidence. Analysis of and conclusions formed from this data are in the author's opinion equally incontrovertible. If life on our world is to continue on its current evolutionary path then pollution and environmental damage resulting from human activity must be curtailed and managed, leading to a reduction to long-term sustainable levels whilst at the same time actively encouraging biodiversity.

In the author's opinion, the work by Meier and Lebot [71] highlighting the potential savings for "standby" equipment as the One Watt Initiative catalysed numerous incentives to reduce wasted electrical power. Requirements to purchase energy-efficient equipment [1] and energy efficient legislation [3] encouraged industry to respond and develop energy efficient goods for the marketplace. These initiatives have culminated with the introduction of generic International Organization for Standardization (ISO) [48, It's all in the name] energy management standards as ISO 50001 [99] to provide a solid base for future evolution in conjunction with equipment monitoring resources and standards such as those provided by the Super-efficient Equipment and Appliance Deployment (SEAD) initiative [95] and implemented using initiatives pioneered by the Clean Energy Ministerial [16].

For any effective energy efficiency strategy to be applied, data acquisition, analysis and monitoring is imperative. ICT is a powerful tool for the measurement and analysis pertaining to energy usage data, as well as formulating strategies to provide optimal flexibility and response of business enterprises. However, energy is needed to provide for the service infrastructure. A number of fairly recent articles describe the necessity for increased energy efficiency within ICT as part of the integrated initiative to redress the issues leading to climate change. Probably the first such article to gain a wide audience was presented as a 2007 report by Gartner, a US-based IT research and advisory company [74]. The report champions the deployment of ICT as a tool for reducing energy consumption and proposes the adoption of a number of existing technologies and practices to result in real financial and energy savings. This work seemed to energise the industry and through the birth of "Green Computing" has led to a considerable voluntary effort within the industry to reduce its own power consumption [76], [62] although the optimum implementation strategy continues to be discussed [103].

As improvements to data centres impacts total ICT power consumption, the next largest consumer within the ICT industry is wireless networking. Numerous works refer to the high energy consumption by cellphone operators which can be 20% or more of the total operating expenditure [62], [30], [73]. This situation can only worsen for cellular communication, since as new standards are released, consumers upgrade their handheld units and cellphone contracts to secure more bandwidth. This trend shows no sign of slowing. A recent publication predicts that the number of contracts will exceed the global population by 2014 [90]. Although a little premature, this event is still predicted to occur shortly [52].

The rapid evolution of wireless communication is well documented by many sources (e.g. [78]). The success of this data transfer methodology is, in the author's opinion, predominantly due to the flexibility of the infrastructure. This is most readily demonstrated by the capacity for and ease of information transfer whilst a user is in motion. However, signals must be consistently transmitted from fixed BSs in order to provide knowledge of the communications infrastructure to the mobile user.

In order to achieve the globalisation which has arisen for mobile communication, telecommunications companies have had to make considerable investments in providing BSs to relay data between the user-held data handsets referred to as the Mobile Station (MS). Initially, the usability of user handsets were restricted by the physical size, weight and power storage capacity of the battery module. However, with the development of lithium ion cells by Goodenough and others (resulting in the 2012 award of the IEEE Medal for Environmental and Safety Technologies [13]), light, high energy density storage cells were available leading to dramatic miniaturisation. Increased portability combined with advances in hardware and software design fuelled the growth of the wireless industry to its current position of a user base of over 6 billion devices and providing 2.4% of the world Gross Domestic Product (GDP); that is to say, wireless communication was responsible for almost $\frac{1}{40}$ of the world's total annual spending [90] and comparable with energy expended by air traffic.

In the following chapter, a review of the literature pertaining to the strategies adopted by the wireless communications industry will be presented. This will be discussed in the context of establishing a benchmark to determine the effectiveness of any solutions proposed in later chapters. Regrettably the chapter begins with a lengthy discussion of the reasons for the choices made therein. This was thought necessary to provide a vehicle to assist the reader's understanding and is not intended to cause irritation.

Chapter 2

Literature Review

> *"... so I went to work. And here I must needs observe, that as reason is the substance and origin of the mathematics, so by stating and squaring everything by reason, and by making the most rational judgment of things, every man may be, in time, master of every mechanic art."*
>
> – Daniel Defoe, *Robinson Crusoe*

As explained in the introduction, it is the author's opinion that the fundamentals of wireless communication must first be outlined and discussed prior to describing the effectiveness of current solutions and providing any strategies for improved energy efficiency. Whilst the literature and standard references utilise advanced mathematics to describe many of the phenomena pertaining to wireless transmission, in this chapter the author has deviated from accepted practice and rather than simply reproduce this material, has instead attempted to provide a reasonably comprehensive overview from a more descriptive perspective containing physical insight using a number of summaries. In commencing this work the author was directed to utilise a specific text, namely [78] as a source of both knowledge and understanding for this project. Having spent some hours alone working with this book as my sole companion, I can now declare that the book provides the author with neither.

In undertaking this review, the author would like to try and break this cycle of confusion. There are many other resources available and whilst the work by Molisch is commendable insofar as it collates much of the pertinent material into a single 'thin' volume, fundamentally, in the author's opinion, as a single source it makes relatively simple concepts harder than they should and makes difficult concepts impossible. The references cited therein do not ameliorate this regrettable situation. Further, to be so confined makes the task of demonstrating that one has performed one's own research more difficult since any critical discussion is likely to read simply as a critique of Molisch's work. In an effort to provide some brevity and understanding for the reader, when faced with writing pages of formulae which are unlikely to be deployed within the current work, or forming lengthy intricate arguments, the author has chosen to direct the interested reader towards popular texts for detailed arguments and present pertinent information as clearly and simply as possible to convey a flavour for the issues under discussion.

Another popular presentation plan favoured by some authors appears to immerse the reader within a dazzling array of diverse networking technologies in an effort to convey the richness of wireless networking development without providing any reference frame from which they may be able to generate an internal structure for this material. In undertaking this work the author has read numerous works and found much of the material difficult to negotiate and to find context for. With hindsight, after investing the effort to assimilate only a small portion of the body of work within the public domain, the author is inclined to suggest that much of the published work seems more intent on justifying its existence than actually educating the reader.

Politics aside, in formulating this review the author has attempted to describe wireless networks in the most generic sense until many of the contributory aspects have been exposed. When describing a scenario in terms of a network, whenever possible the author has elected to try and utilise the properties of a stand-alone IEEE 802.11 wireless network. The reasoning behind a single preferred choice is that each wireless infrastructure utilises different ideas, heuristics, protocols and combinations of components to provide a working ***product***. Each product works well, but could it work better and as engineers do we possess the knowledge, understanding and maturity to identify how to accomplish this? Based on the literature reviewed for this work, the author is regrettably prepared to stand up and be counted in the "Nay" vote.

In wired networks, the Medium Access Control (MAC) protocol war was won by TCP/IP running over Ethernet due to its high speed, ease of installation, superior reliability and efficient handling of contention. Commercially, it offered the best cost-effective solution which made it adaptable to evolving technology. Now that the war is over, increasingly faster connections are made in the absence of contention in wired networks. This is not a straightforward analogy which can transfer easily to wireless. Radiation leaks away from its source and reduced interference is frequently a result of agreement between users of the shared medium rather than any control one may exert over media properties as in a wired network. Throughout the history of ICT, so there has been a convergence of products based upon commercial pressures rather than on products' ability to adapt and address the problems they were originally intended to solve. The author simply wants to try and understand and in writing this report, to help others to do the same.

When presenting their information, authors appear obsessed with incorporating their own individual set of wireless abbreviations in their work. Having to constantly turn to the glossary section comprising several pages of acronyms to understand the most mundane sentence is something the author finds frustrating, time consuming and annoying since it breaks one's train of thought. As one reads more and becomes more familiar one becomes more inclined to guess abbreviations rather then consult the glossary, almost as if undertaking some cryptic puzzle to be able to congratulate oneself on one's cleverness if the guess is correct. In the author's opinion this practice is quite addictive and potentially dangerous since an unchecked incorrect interpretation of an abbreviation can lead to a lack of understanding that is neither easy to identify or remedy. Overall, it seems that for the majority of published work within wireless networking, the intended target audience comprises the very people who don't need to read the work at all! The lack of a standard set of abbreviations in a topic so obsessed with standardisation (which presumably is why there are so many) seems utterly incomprehensible but clearly remains a work in progress. Where possible the author will attempt to retain the terminology utilised within the IEEE 802.11 standard [10, Chapter 3: definitions, acronyms and abbreviations].

For much of this section, when a data transfer over a Wireless Medium (WM) is taking place this will be described as being between a {stationary node, STAtion (STA) or BS} and an {(optionally) mobile node, STA or MS}. There are a number or conceptual difficulties associated with this that have not escaped the author's attention, not least of the possibility of a mobile BS and stationary MS. Whilst this particular instance may be circumvented by considering the two nodes in isolation and changing the reference frame, other difficulties cannot be dispensed with as easily. Specifically, the principal terminology of wireless networks originates from two disparate sources. Data communications are described by the IEEE 802.11 standard [10] in terms of a cluster of STAs or nodes forming a Basic Service Set (BSS). Should the Access Point (AP) within a BSS be connected to other BSSs via an integrated Local Area Network (LAN) such that the assembly appears as a single BSS to the Logical Link Control (LLC) of any given STA, then this assembly may be considered an Extended Service

Set (ESS). Telecommunications appear to refer to the central ‘stationary’ node within a cell as the BS servicing the other ‘mobile’ nodes as MS.

Networking applications are traditionally viewed in terms of the ISO/Open Standards Interconnect (OSI) seven layer model, for which the reader is referred to the introductory chapters of any of the more established networking texts [97], [38], [91] if a reminder is needed.

The author is of the opinion that in order to understand a data network from the bottom up, one must first appreciate the limitations of its physical layer. All wireless communication utilises analog transmission over a shared communication WM. As such, any wireless networking environment must be considered a physical bus network. Logically a wireless network may be viewed as either a star network (Characterised by having a central device that defines common sets of connection parameters such as range, communications protocols supported, procedure for loss of the central device etc.) or a mesh network (Point to Point, Peer to Peer, ad-hoc or sensor network, e.g. a wireless sensor network) connecting devices accordingly.

The first section of this chapter outlines the physical environment supporting wireless networks and provides a flavour of the specific issues relating to reliable wireless transmission. The next section introduces terminology specific to wireless communication and begins to address the problems introduced by transporting a signal through its shared environment. This is followed by a brief outline of the quantities usually examined by networking specialists to assess the performance of a network link and the allowances that must be made in order to maintain a working connection using a link budget (Section 2.2).

Since the beginning of wireless, engineers have developed techniques to improve the quality and reliability of connections using a strategy termed diversity in section 2.3. These are summarised in a self-contained section that should provide some comfort to the reader insofar as offering hope that a reliable high capacity link is still attainable even when faced with devastating impediments to the contrary.

One of the currently unmentioned challenges in this review is to find strategies for efficiently sharing the wireless medium which can be addressed in section 2.4. Once the essential technologies utilised in current wireless transmission have been presented, an overview of wireless standards with specific reference to power management will be made in section 2.5, finishing with some general topics relating to efficient, “environmentally friendly” energy sources for wireless transmission in section 2.6. The chapter concludes with a summary and discussion of the information presented. The author hopes that this will provide a platform to present the work programme undertaken in this work and examine the challenges of energy efficient wireless communication explored in following chapters.

2.1 Wireless transmission fundamentals

Whilst the label, "wireless" can conjure images of diverse information transfer methods employed throughout history, the current work will restrict itself to electromagnetic radiation lying between the "radio frequency" portion of the spectrum, specifically with frequencies in the range 3 kHz - 300 GHz. All wireless transmissions comprise analog signal transmissions of the baseband data stream superposed on a carrier wave. An overview of the fundamentals are presented in the first subsection on modulation. Many of the issues presently limiting wireless data transmission rates occur as a result of the interaction of electromagnetic radiation with matter, which will be reviewed briefly in the second subsection. This constitutes a deviation from the texts reviewed for this work which prefer to first describe the history of wireless or present an overview of wireless deployment. The author has chosen this particular approach based on the opinion that without an appreciation of the issues affecting data transmission, the motivation for adopting particular higher level strategies for improving the quality of a wireless link will not present itself.

2.1.1 Modulation

Modulation must therefore utilise at least one of the characteristics amplitude, frequency, phase of the carrier wave. The digital nature of the data is therefore coded to discrete levels using combinations of {Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK) or Phase Shift Keying (PSK)} respectively. The reader is directed to standard texts (see for example [91, Section 5.2], [38, Chapter 5], [44, Chapter 5]) should they need to reacquaint themselves with the basic ideas.

Molisch deems modulation important enough to merit a chapter of its own [78, Chapter 11] without it seems the need to recognise the contribution from standard texts to fundamental modulation ideas or basic signal processing. Whilst the material presented here appears insightful the author confesses to some confusion behind the driving force. Early within the text, Molisch has stated that RF electronic circuitry is highly specialised and beyond the scope of the content. After having undertaken some background reading into this and other peripheral topics whilst researching the background for the current work, the author concurs completely. However, in his modulation chapter, Molisch firstly expresses the desire for waveforms to possess a profile more suited to non-linear high energy efficiency amplification. He later goes further to carefully expound the need for a filtered basis pulse from a 'standard' square wave to that of a Nyquist pulse so that once the signal is encoded it will convey a faster reduction in power away from the central frequency and so reduce Inter Symbol Interference (ISI).

Whilst the author is in complete agreement with the sentiments, it is the author's opinion these matters are an extension of the electronic circuit design principles and whilst interesting, do not and should not be a serious consideration of the target audience of this work based on the following reasoning:

Periodic electrical signals may be decomposed into a series of sine and cosine waves according to Fourier's theorem. On this basis the artefacts occurring as a result of generating a 'pure' square edge can be readily viewed and appreciated as Gibbs' phenomenon and further extended to understand the constraints levied on the upper frequency capability of electronic equipment from signal generators and beyond (see for example, [93, Programme 6], [55, Chapter 10]). Extension of these ideas lead naturally to the realisation of the imperceptible reality of a square wave because one simply cannot produce or monitor signals with infinite frequency! (The other more general realisation is that discontinuities of any form frequently lead to unwanted harmonics and so should be avoided whenever possible.) In practice, 'square' waves utilised for clocking in digital circuitry are generated using something similar to an astable multivibrator circuit: effectively the 'vertical' portion of the square wave is in fact derived from the clipping of a high gain transistor output and is a curve having a very high magnitude gradient.

The author's argument then is that the basis pulses could simply be generated at source within the circuitry and abstracted as such to the reader. Having now established that the voltages generated within circuitry do not need to be and indeed are fundamentally not perfectly angular, the notion that one may produce any form of signal within a circuit for processing is readily apparent: the additional complications and expense associated with producing linear circuitry is not necessary in the author's opinion. To try and put this into perspective, consider the following over-simplistic example. To generate a monotonic linearly increasing straight line output with increasing time, one may either input a linearly increasing $V_{\text{in}} = kt$ signal into a linear voltage gain amplifier $V_{\text{out}} = GV_{\text{in}} = Gkt$ or a $V = \sqrt{kt}$ signal into a quadratic (i.e. non linear) voltage gain amplifier $V_{\text{out}} = G(|V_{\text{in}}|)^2 = G(\sqrt{kt})^2 = Gkt$. The author recognises and accepts that this is a gross oversimplification of the underlying processes but remains of the strong opinion that the principle of the example should be verifiable as a proof of concept at least following simulation with a tool such as Cadence Virtuoso [11] and as such this can be abstracted and mitigated from consideration of modulation requirements. Clearly the important feature here should be the production of radio signal waveforms conveying the maximal amount of information from the lowest possible energy requirement to provide a received signal that may be processed at the receiver to recover the original data.

Within a wireless transmission, the modulation signal should also contain intrinsic timing information to avoid further encoding or the provision of a separate data channel which may be treated differently by the propagation path, making data recovery even more challenging. This will likely be derived from the properties of the carrier wave itself. One interesting modulation format which results from the lengthy reinterpretation by Molisch is that of Minimum Shift Keying (MSK), also known as Continuous Phase (CP)-FSK. Here, the spurious features generated from discontinuities in signal output tolerated within other modulation schemes are addressed by ensuring a continuous phase occurs with a frequency-modulated signal. This is accomplished by having only two single bit symbols represented unique frequencies which perform complete cycles within the symbol period and changing the frequency at carrier wave zero crossing points. Gaussian Minimum Shift Keying (GMSK) provides further benefit compared with MSK since it utilises smoothed basis pulses resulting in greater frequency drop-off away from the central transmission frequency to give better spectral efficiency.

When implementing a connection, the modulation format used will depend upon the amount of noise within the system. In extreme cases, high noise resistance is conveyed by avoiding amplitude modulation since these signals will be affected by noise superposition to a greater amount than constant amplitude modulation formats. This will however reduce the modulation order, decrease the number of bits transferred per signal and so spectral efficiency will decrease. Lower extraneous noise levels mean that higher order modulation can be leveraged with increased spectral efficiency. Molisch [78, Table 11.1, p219] summarises the spectral efficiency conveyed by different modulation schemes in a table which is reproduced below to assist the reader. Note the difference between the quoted spectral efficiencies of Binary Phase Shift Keying (BPSK) and MSK. BPSK can incur discontinuities at transitions resulting in a reduced spectral efficiency.

Table 2.1: Spectral efficiency for different modulation schemes, after Molisch [78, Table 11.1, p219]

Modulation method	Spectral efficiency for 90% total energy (/ (bit/s/Hz))	Spectral efficiency for 99% total energy (/ (bit/s/Hz))
BPSK	0.59	0.05
BAM ($\alpha = 0.5$)[a]	1.02	0.79
QPSK, OQPSK	1.18	0.10
MSK	1.29	0.85
GMSK ($B_{\mathrm{G}} = 0.5$)[b]	1.45	0.97
QAM ($\alpha = 0.5$)[a]	2.04	1.58

[a] Roll-off factor: a parameter representing the steepness of the spectral decay.
[b] Bandwidth of the Gaussian filter as a proportion of bandwidth corresponding to 99% overall energy.

Practically, wireless devices are configured with rules for selecting a particular Modulation Coding Scheme (MCS) according to the noise and responsiveness of the network at a given time [7, Chapter 4]. The ability to adapt the modulation scheme and the coding rate of the error correction according to the quality of the radio link is known as link adaptation.

2.1.2 Radio fundamentals

Radio frequency waves are generated by passing an oscillating current through a conductor acting as an antenna. The radiation pattern produced is strongly dependent upon the antenna geometry and all but the most general observations are beyond the scope of this work. As with all radiated signals, the magnitude of the radiated power follows an inverse square law in free space, as described using Friis' law:

$$P_r = P_t G_r G_t \left(\frac{\lambda}{4\pi d} \right)^2 \tag{2.1}$$

where P_r and P_t are the received and transmitted powers respectively, G_r and G_t are the power gains for receiver and transmitter antennae and circuits, d is the separation between transmitter and receiver and λ is the signal wavelength.

Note that this expression includes the term for the reciprocal of Free Space Path Loss (FSPL) at the far right:

$$FSPL = \left(\frac{4\pi d}{\lambda}\right)^n \tag{2.2}$$

where n is given as the generic free space loss exponent. Whilst derived from first principles and typically quoted as $n = 2$, it is however possible to vary this parameter to reflect path loss in different environments using a simple log-distance path loss model [84, Section 4.9.1]. n can vary anywhere between $1.6 \leq n \leq 1.8$ for internal line-of-sight situations and can range from $3.5 \leq n \leq 4.5$ for non line-of-sight (for example urban areas or within a stadium) [78, Section 3.2.1]. Rappaport extends the exponent up to $n \approx 6$ for extreme cases within a building but this was not confirmed by the measurements conducted by Miranda et al. [75].

As with all parts of the electromagnetic spectrum, radio waves comprise an electric field oscillating at right angles to a magnetic field; both of which are perpendicular to the direction of energy transfer. In some instances, the oscillations of two waves may differ in the direction of their electric fields. In this situation the waves are said to be polarised. Although frequently ignored, polarisation occurs in waves far more frequently than one might imagine, from the production of the waves to the effect of say reflection from a material surface. The effect is not generally considered since it can be difficult to extract the different polarised radiation. However, as explained later in this chapter, extracted polarisation of radio waves can be undertaken to improve signal quality in some situations.

As stated previously, a wireless signal travelling through the shared atmosphere are more susceptible to interference than the directional media encountered within wired networks. The sources of interference are outlined in the subsections below.

2.1.2.1 Physical obstacles

Wireless signals can be strongly absorbed by large thick insulating objects (such as buildings) or reflected by highly conducting objects (such as large bodies of water or steel walls). Should the line of sight between a BS and MS be obscured by such an object, then the received power will decrease considerably. This phenomenon is known as large-scale fading.

Less strongly absorbing dielectric materials will allow the electromagnetic radiation to penetrate through. However, in passing between media different proportions of the wave will be reflected, refracted and absorbed. Reflected signals can lead to the signal being sent to the receiver by more than a single path, most likely with a different path length. Effects concomitant with this phenomenon are referred to in the literature as multipath. The differing path length leads to differences in signal propagation time, producing a time dispersion in the received signal. Multipath effects contribute substantially to the degradation of wireless transmission, as the following discussions will explain.

Reflection from obstacles can produce undesirable interference effects. Ideally, to mitigate first order interference effects from reflection, the region within the first Fresnel zone should be kept clear of obstacles (This is used phenomenologically to determine where to site permanent wireless antennae or whether to remove obstacles such as trees). Fresnel zones represent a family of ellipsoidal surfaces with transmitter and receiver at the principal foci. One of the properties of an ellipsoid rotated about the line of its principal focus is that the distance between the foci and the surface is constant for all points. Therefore, any reflections at the surface of a given Fresnel zone will incur the same path length. For the nth zone, the additional path length from the line of sight is given as $n\lambda/2$ which corresponds to a phase change of π^c, as given by Molisch [78, Section 4.3.1]. In this reference, Molisch also recounts the criterion for the breakpoint deviation from Friis' law in terms of the intersection of the first Fresnel zone and the ground.

Reflection may also provide enhanced data transmission between long wall-like obstacles, for example in a corridor or a street. These barriers guide the wave and so the process is known as waveguiding. However, unlike waveguides utilised in other situations (for example, conducting metal waveguides for microwave transmission between decks in ships or say light passing through optical fibre), wireless waveguides are subject to greater power loss, both from the partially absorbing wall medium as well as from any irregular surface texture.

Refraction can be employed to increase the transmission distance of wireless signals. The refractive index of the atmosphere decreases with increasing height above Earth's surface, causing the direction of a radio transmission to be "bent" back towards the Earth. However, for shorter range transmission it is conceivable that changing the direction of a transmission may cause the transmission path to be affected. No reference has been found for dispersive effects occuring as a result of refraction for radio frequency transmission so this is presumed

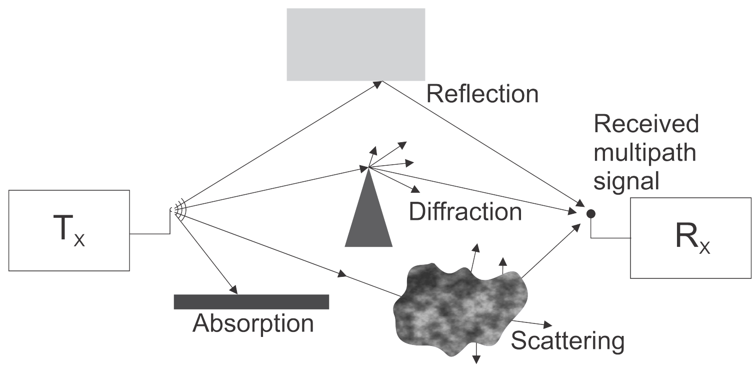

Figure 2.1: Schematic representation of interactions between wireless signals and the environment to give a multipath received signal

insignificant.

Further disruption of the signal may occur at the edges of an obstruction. Where the size of an object is comparable with the wavelength of radiation, diffraction can occur. Diffraction effects cause slight changes the wavelength of the radiation, but the phenomenon has a more marked effect on the direction of transmission. This phenomenon may ameliorate transmission around obstructions but may also provide further opportunity for multipath transmission leading to increased interference from multipath effects. Moreover, diffraction from antennae produce untoward effects upon wave propagation. For example, Friis' transmission law is only valid for far field, that is to say, distances exceeding the Rayleigh distance [78, p48].

Some obstacles may not be fixed: for example as an electromagnetic wave passes through the atmosphere, it will not only be partially absorbed by atmospheric gases but will also be scattered by rain [47], snow and fog. This scattering does not occur in the same cooperative manner encountered with diffracted signals, but rather leads to an uncoordinated scattering and dispersal of the incident radiation which is modelled probabalistically.

These phenomena are illustrated schematically in Figure 2.1. For more information, the reader is directed to the treatments given in standard wireless texts ([78, Chapter 4], [32, Chapter 6-8]).

2.1.2.2 Noise and interference: the disruption of radio frequency signals

Interference refers to a phenomenon which occurs when waves meet. Typically, at a given location in a given instant, the total resultant wave displacement may be expressed as the algebraic vector sum of the individual wave displacements. If a given information transmission comprises a particular wave pattern and this transmission passed from transmitter to receiver via two paths, then these wave patterns would interfere at the receiver to produce a resultant wave having a different pattern to that originally transmitted. The amount of interference experienced would depend upon the relative signal amplitudes: destructive interference will be greater for signals of similar amplitude. Nevertheless, interference will cause the raw information received to differ from that originally transmitted.

By comparison, noise can be simply considered to comprise unwanted changes to a signal, possibly originating from unknown sources. Noise comprises irregular fluctuations which are received with a received signal but were not an intentional transmitted part of the signal and act to obscure it. Practically noise cannot be eliminated: its effect can only be reduced. For this reason considerable effort has been expended in defining the behaviour of noise originating from particular sources. The most common sources of wireless noise are summarised briefly below to provide a quick reference for the reader.

Johnson-Nyquist noise occurs within electrical conductors as a result of random thermal motion of charge carriers (most commonly electrons) and so will be produced by the circuitry transmitting and receiving the wireless signal. This type of noise has is characterised by having effectively constant average power over all frequencies (constant power spectral density). The power spectral density increases with increasing temperature and so is also referred to as thermal noise. Subjecting the noise level to statistical analysis shows that the amplitude variation can be usually modelled as exhibiting Gaussian behavior centred on zero displacement. The most common model for thermal noise is frequently described in the literature as Additive White Gaussian Noise (AWGN).

Impulse noise is encountered as an intermittant, high amplitude transmission from natural (such as lightning discharges or sunspot activity) or artificial (such as power cable surges or heavy electrical switching encountered with say an arc welder) sources. The size of the disturbance is frequently equivalent to the magnitude of the size of the desired signal or exceeds it but is typically of short duration compared to the data transmission sequence, which is fortunate since the signal is not easily recoverable and often requires retransmission.

1/f or flicker noise is generated in semiconductor devices at what is thought to be Si/SiO_2 interfaces. As the name suggests, the noise diminishes with increasing frequency but is particularly pronounced at radio frequencies within low power MOSFET devices. This type of noise can usually be reduced at source by electronic filtering.

Although the terms interference and noise are typically interchanged quite easily, it is considered useful here to make a distinction between them. Interference is described here as the incorporation of unwanted signal into the useful signal. Again, there are many sources of interference, the most important of which for radio frequency transmissions are briefly summarised below.

Inter Symbol Interference (ISI) occurs when multipath propagation occurs, most frequently as a result of the transmitted wave travelling via two or more paths say by reflecting from obstacles. This occurs since the information transmitted within the signal travels at a constant speed (the speed of light) and so the time needed to cover different path lengths will also differ. The transmitted data is referred to as a symbol, which may comprise one or more bits of data. Upon receipt of the signal, any time delay will result in a broadening of the information contained within the signal, ultimately causing degradation by "blurring" the data. Overlaying one symbol with another (whether it is the same symbol or a different one) in this way is often referred to as ISI.

Co-Channel Interference (CCI) or crosstalk is caused by two transmitters in close proximity radiating different signal information at the same frequency. A wireless channel is used to describe the given frequency range within the spectrum being used for the wireless communication. Typically Co-Channel Interference (CCI) is most easily avoided by either negotiating to ensure that only a single source will transmit data within a given time slot, or if this cannot be undertaken, by designing an infrastructure that will not incur this problem (increase transmitter separation, frequency planning to ensure different transmission frequencies). However, as will be seen in 2.5.1, this issue is more frequently encountered when maximising usage of the available spectrum.

Adjacent Channel Interference (ACI) arises due to two transmitters radiating at comparable frequencies. To increase the number of networks exchanging data, transmitters will utilise different frequency bands (or channels). Despite standards retaining unused frequency ranges (guard bands) between channels, should a transmission encounter any extraneous power leaking from an adjacent channel then the transmission will suffer Adjacent Channel Interference (ACI).

Typically, the receiver must attempt to recover data from a transmitted signal after it has been subjected to modification by the physical environment and any spurious comparable frequencies present between BS and MS. However, routinely monitoring a "noisy" signal does not easily demonstrate a pattern with properties inherent in the original transmission. Put simply, interference can lead to a noisy signal but noise cannot be processed as a signal interference.

The strongest sources of Radio Frequency (RF) interference are found from devices that operate at similar frequencies. To illustrate, wireless LANs based upon the IEEE 802.11 b, g, n standards operate at least partially in the 2.4 GHz band reserved for general-purpose Industrial Scientific and Medical (ISM) purposes (802.11n can also optionally operate in the 5 GHz frequency band). (For this frequency, from $\lambda = c/\nu$, the wavelength is $3 \times 10^8 ms^{-1}/2.4 \times 10^9 Hz = 0.125m$.) This band is also utilised by bluetooth devices and cordless telephones, which can occasionally lead to interference [7, Chapter 13 - Configuring Cisco CleanAir]. Additional sources are produced by other nodes participating within wireless networks. To illustrate, consider devices communicating using the ISM band. A LAN will operate in one of the twelve different frequency channel ranges offered within the band. These transmissions will not only be affected by co-channel interference from devices within the same LAN and nearby LANs using that same frequency, but also by ACI from LANs operating at neighbouring frequencies [60]. ISI will only occur if there are obstacles such as walls that would provide multiple paths between transmitter and receiver.

2.1.3 Summary

The preceding topics have presented modulation schemes developed to mitigate signal disruption by the many different processes present during the transmission of a radio frequency signal. Due to the vast differences in situation associated with each effect, it is difficult for even the most knowledgeable and experienced individuals to assess a given situation and deduce the most important effect occuring. Consider for example the transmission of radio frequency signals over an urban landscape. The principal obstacles are buildings which may be approximated as having a well-defined cuboid shape. However, the construction materials employed affects the relative amount of absorbed, reflected and transmitted radiation as well as the extent of any polarisation: the texture of the material has strong effects on the amount of scattering and the edges of the cuboid provide diffraction effects. Any multipath transmissions must be considered with regard to the specific location of the receiver. Furthermore, for a single wireless transmission between BS and MS in other than ideal free space conditons, the effect of multipath would need to be considered repeatedly. Models have been constructed using ray tracing techniques to investigate these effects [78, Section 7.5], [44, Section 2.4]. However, these models are not always successful, leading to the deployment of phenomenological models such as the for urban landscapes (see for example [78, Chapter 7: Further reading]).

To accurately predict the principal effects acting to diminish the reception quality of a transmitted signal is clearly nontrivial. However, to provide an effective service to mobile users, wireless system providers must not only understand and manipulate these effects but also counter any disruption introduced into signal reception.

Clearly, in order for data to be recovered from a received signal, the transmitted information must have power exceeding that of the degradation occuring during transit. Friis' law (2.1) shows that the only ways to increase received power for a free space transmission are to increase transmitted power and component gain. Increasing gain requires comprimises to be made, since arbitrary increases in power can introduce additional noise into the system. The alternatives are to carefully manufacture or select the amplifiers to have low noise (which increases cost), to decrease data rate (which is undesirable from a performance perspective) or to utilise additional channel bandwidth. The latter measure is both difficult and expensive since utilised frequencies can either be freely available and contain other signals which must be extracted from the received signal (such as in the ISM band) or can be exclusively leased from Government (see for example [18]). However, recently some manufacturers have begun to produce components which allow mobile devices to utilise other portions of the wireless spectrum [57].

A more sensible alternate strategy might be to incorporate mechanisms to recover the transmitted information more effectively. In pursuit of this objective, as wireless communications have evolved, so has the need to obtain a quantitative measure of the amount and type of noise and interference occuring within a given signal under actual deployment conditions. This may allow noise reduction techniques appropriate to the situation to be applied but also allows the effectiveness of any corrective adjustments to be monitored in order to optimise the quality of the data links provided. These measurements will be outlined briefly within the following subsection.

2.2 Parameterisation of the quality of a wireless signal

Many of the terms used to describe the properties of radio reception may sound familiar to the layman. These and other less familar terms will be discussed below in terms of the processes outlined in the previous subsection. It is the author's most fervent desire to offer safe passage through the literature to the reader whilst they encounter this dazzling array of technical terms which is by no means exhaustive. The presentation below attempts to provide context whereby the reader may start to manage the complexity inherent within this rapidly expanding field of human endeavour. Indeed, in both planning and writing this section, the author has frequently felt overwhelmed by the sheer enormity of the problem faced in establishing a reliable wireless link. The numerous threads are drawn together with the summary which serves to establish a foundation upon which many highly intelligent and capable people have constructed the reliable wireless communications framework enjoyed by so many today.

2.2.1 Gain

Gain is an ambiguous and potentially confusing term used to measure some output from a component or device compared to the amount of input i.e. gain = output / input. There are different possible variants of gain, depending on the property of the signal being measured. The most common device for which gain is measured is an amplifier which increases the power output of a signal using additional power from an external supply. All amplifiers (or active devices) provide a power gain, however the mechanism by which this is accomplished can be to provide increases in the voltage (voltage gain), current (current gain) or both. The confusion is compounded since as a dimensionless parameter there are no units which may be used to indicate which property has been measured.

In radio frequency applications the literature most frequently refers to power gain:

$$Power\ gain\ G = \frac{Output\ Power\ P_{\mathrm{out}}}{Input\ Power\ P_{\mathrm{in}}} \tag{2.3}$$

Unless specifically stated, all subsequent references to gain in this work will refer to power gain.

Gain values vary over enormous ranges and so are frequently cited using the decibel (dB) scale:

$$Power\ gain\ G_{\mathrm{dB}} = 10\ \log_{10}(G) \tag{2.4}$$

For wireless applications, gain is realised from the RF power amplifier. The power amplifier has the highest power consumption requirement of all the components needed to provide a wireless network and probably embodies the single most significant source of technological challenge of all wireless engineering components. Not only is the electronic circuitry a source of flicker noise, but the high power consumption increases local temperatures leading to greater Johnson-Nyquist noise. Furthermore, the high temperatures and high voltages associated with transmission not only place exacting demands upon the electronic component themselves but also the component layout, the conducting tracks connecting them and the electrically insulating substrate upon which the devices are assembled. This is further compounded by the need for controlled thermal management to conduct excess heat away from the components. Operationally, other factors, including but not limited to, cost, weight, efficiency, reliability, signal distortion, ease of design, high temperature operation, frequency response and operating bandwidth describe some of the considerations facing electrical engineers as they strive to provide and select usable products for a particular intended deployment. Even the most basic exposition of the requirements of these components is beyond the scope of this work. The interested reader is directed towards specialist texts (e.g. [45]) for further information.

2.2.2 Antennae

Radiated radio frequency signals must be transmitted and received using an antenna. A directional antenna can be used to concentrate the supplied incident power into radio waves along particular directions compared to an isotropic antenna which radiates equally in all directions (at least theoretically). For a wireless network, transmitting in some directions would be considered a waste of power. For example, for a terrestrial application, any transmission in a vertical direction towards or away from the Earth will not easily reach recipients.

Antenna design is a highly specialised discipline beyond the scope of this work, but as the essential component for wireless transmission, the properties of different antennae is a fundamental consideration in any radio frequency wireless system and so is briefly considered here. One measure of performance for an antenna is the antenna power gain G:

$$Antenna\ power\ gain = \frac{Power\ received\ from\ antenna}{Power\ received\ from\ an\ isotropic\ radiator} \tag{2.5}$$

Where both radiators have the same incident power. Unlike the gain from an amplifier introduced above (2.3), power is not added by the antenna: rather the available power is distributed to provide higher measured power in particular directions at the expense of reduced power in others. The radiation profile from an antenna is usually shown as a radial pattern with long range lobes in particular direction. Antenna gain is therefore a passive phenomenon. Usually literature discussions of antenna power also introduce the concept of antenna efficiency at the same time:

$$Antenna\ efficiency\ \eta = \frac{Radiated\ power}{Incident\ power} \tag{2.6}$$

However, in directing emissions, the radiated power profile becomes dependent upon emission direction and is difficult to parameterise effectively. Popular parameters used to describe the emitted power profile are summaried below.

- Far field directive gain $D(\theta,\ \phi)$:

  $$D(\theta,\ \phi) = \frac{Power\ radiated\ per\ unit\ solid\ angle\ in\ direction\ \Omega(\theta,\ \phi)}{Average\ power\ radiated\ per\ unit\ solid\ angle} \tag{2.7}$$

  (where θ and ϕ are the polar and azimuthal angles from spherical polar coordinates).

- The direction of maximum directive gain, known as the boresight direction.
- The "width" of the pattern is described by the beam width. This comprises angular values for which the directive gain (or radiated power) is a proportion of the gain (or radiated power) measured along the boresight direction. Probably the most common measure is the -3 dB beam width, effectively corresponding to the half power value. The seminal work by Nelson for a circular dish antennae [79] shows that the half power beamwidth decreases with decreasing wavelength and increasing diameter. Interestingly, the work also derives an expression showing that the antenna gain is inversely proportional to the square of the half beam width.
- Antenna bandwidth describes the range of frequencies for which an antenna operates effectively. For all practical purposes, this should correspond closely to the frequency range transmitted by the antenna.

When deploying antennae in a radio network, one of the practical optimising adjustments is to tilt an antenna such that the boresight direction is aligned below the horizontal plane [50, Section 4.4.3.5], [78, Section 9.3.3]. This has the advantages of reducing path loss, delay spread and system interference, enabling the antenna to operate with lower transmitted power. Also, by reducing the amount of low power waves escaping to any nearby networks, CCI is reduced, thereby increasing network performance.

The interested reader is referred to specialised antenna (e.g. [42,61]) or standard wireless texts (e.g. [50], [78, Chapter 9], [32, Chapter 4]) for additional information on this topic.

2.2.3 Fading

Fading is a term used to indicate a diminished received signal strength. Extreme or deep fading may lead to a prolonged loss of communication and would likely require external remedial action so will not be considered further here. There are many different causes and classifications of shallow fading, some which are outlined briefly below to illustrate the inherent transmission problems.

2.2.3.1 Large scale and Small scale fading

Large scale fading, or 'shadowing', occurs as a result of the line of sight between BS and MS being blocked by say a large body. In this instance, the MS would need to move a large distance (compared to the transmission wavelength) to re-establish line of sight conditions. This would likely result in a loss of synchronization between the stations for a short time delay leading to a complete loss of data exchange with increasing delay.

By comparison, small scale fading is indicative of effects which arise from small changes in signal strength with small changes in distance leading to a change in transit time. To illustrate the essential characteristics for distance, consider ISM waves with wavelength 0.125 m. When one considers that a path difference of $\lambda/2$ between two such waves can cause a result in a transition from constructive to destructive interference, then should a 2.4 GHz ISM wave incur multipath wave transmission, then only a 62.5 mm path difference could change the received signal power from a maximum to a minimum. This path length difference results in a broadening of the received signal in time. Parametrically, the standard deviation of this time broadening is termed the delay spread.

For the remainder of this discussion, all fading effects will be considered to be acting on a small scale. Small scale fading can also occur due to Doppler broadening as explained in the following section.

2.2.3.2 Doppler spread and Coherence time

The Doppler effect for sound is well-known as a result of its frequent occurrence in industrial society. The change in frequency (or pitch) produced between a moving source and observer can be encountered in numerous situations, most notably from the siren of a passing emergency vehicle or perhaps a vehicular speed trap. A similar situation is encountered with wireless transmissions subject to moving transmitter, receiver or surrounding objects. The ensuing Doppler shift becomes larger as the velocity of the interacting object increases and reduces as the cosine of the angle between transmission and movement directions. The standard deviation of Doppler shifts in a given channel is known as the Doppler spread. The frequency modulation resulting from interference introduces small scale fading into the transmission.

Coherence time is a complementary measure to the Doppler spread. Coherence time is the time for which the channel may be considered constant with regard to shifts arising from the Doppler effect. Specifically:

$$Coherence\ time\ T_c \approx \frac{1}{Maximum\ Doppler\ spread\ f_D} \tag{2.8}$$

2.2.3.3 Fast and Slow fading

From the preceding discussions, it should now be apparent that the amplitude of a transmission can change with time as well as displacement. The distinction between fast and slow fading depends upon how rapidly the transmitted signal changes compared to the rate of change of the channel itself. Fast fading will occur if the input channel response changes rapidly within the duration of a signal, otherwise the fading may be considered to be slow.

To gain some insight into their causes, consider a line of sight transmission between a BS and MS. If this line of sight were to be obscured by a moving object (e.g. a lorry with high metal sides) and the object were moving slowly to give a low Doppler spread, the reduced amplitude of the received signal would persist for a longer time resulting in slow fading. By extension, if the object were moving more rapidly, the shadowed signal received would have been subject to a brief period of fast fading.

Clearly this simple illustration is not the only cause of these effects: for example multipath propagation can cause diminished received signal amplitude as both fast and slow fading. Fast fading can be produced from atmospheric effects such as tropospheric scintillation, where turbulent air currents can produce small (1 mm - 100 m) short-lived regions of slightly changing refractive index. Over longer transmission distances (such as those present in WiMax / HiperMAN networks) the small disturbances from tropospheric scintillation can lead to significant losses of signal strength.

2.2.3.4 Flat and frequency selective fading

Flat fading is considered to be independent of frequency. One source of frequency selective fading occurs as a result of symbol broadening occurring from multipath interference effects. Specifically, the delay dispersion resulting from the different path lengths encountered lead to a frequency dependence within the data received. This manifests phase changes and concomitant amplitude variation in the signal which is dependent upon the measured frequency. Other sources may originate from wavelength variations occuring say as a result of diffraction from an obstacle. The general consensus in the literature seems to be that frequency selective fading can be avoided if the signal transmission is restricted to the portion of the physical channel which does not incur any frequency-dependent effects; i.e. flat fading occurs for the the signal bandwidth is smaller than the coherence bandwidth; conversely the delay spread is smaller than the symbol period. Suffice it to say, any real channel can be considered to be frequency selective if analysed over a large enough bandwidth.

2.2.3.5 Temporal dependancies: level crossing rate, average fade duration, mean channel delay and channel delay spread

Probably the most commonly-asked question by people facing any kind of disruption is, “How long will it last for?”. Similarly, with a wireless link, some insight into the duration of fading is imperative for most if not all scenarios. This insight is provided by statistical measures from the measured received power and are described below:

Level crossing rate provides the number of instances where the magnitude of the received power rises (power increasing) or falls (power decreasing) with respect to some arbitrary set level I per unit time. This gives a measure of the frequency of fades occurring for a given environment.

The average fade duraton specifies an average time for which the received power exceeds some arbitrary set level I. If the fade duration is too long, then synchronisation may be lost leading to further errors in data transmission, or in extreme cases the connection may time out completely.

From the preceding explanations in this section, it is apparent that both these measures provide a measure of the amount of multipath within the received signal. Intuitively one might like to determine a probability distribution for the fade duration and so estimate the likelihood of a communication loss under particular circumstances. Due to the complexity of the environment undoubtedly a simplified model for fading would be needed.

Whilst many statistical models of fading have been proposed, most notably by Rice [85] for different circumstances, it is difficult to routinely compare and contrast the effects of these models in generic situations. The interested reader is directed to standard texts which discuss the more common models in detail and provide references to others [78], [44], [32]. The properties of the most commonly utilised fading models in the literature are given below:

Rayleigh fading

Multiple objects cause scattering leading to many (ideally infinite) signal multipaths with no direct line of sight path. Contributions from reflection and scattering combine to produce a resultant intensity with a Rayleigh fading distribution at the receiver. Mathematically simulated by combining two zero mean, non-correlated time independent random variables using the Central Limit Theorem and then square rooting the ensuing distribution to produce a Rayleigh fading distribution.

Rician fading
: Displays similar characteristics to the Rayleigh distribution but having a dominant stationary non-fading (most commonly) line of sight signal in addition to multipath reflection and scattering.

Lognormal fading
: Large scale variations caused by shadowing from obstacles. Multiplicative intensity effects from obstacles effectively lead to an example of the Central Limit Theorem in the logarithmic domain, producing a Gaussian distribution when signal strength is measured in dB.

The power delay profile measures the power of the transmitted multipath signal as arrives at its receiver as a function of the multipath delay incurred. From this data, the mean channel delay or average fade duration $\bar{\tau}$ provides a measure of the average multipath delay present. This in turn depends upon the speed of a mobile station, decreasing with increasing Doppler frequency. The channel delay spread σ_τ represents the Root Mean Square (RMS) or standard deviation multipath delay from the mean. The inverse of this quantity is related to a quantity termed the coherence bandwidth B_C. Coherence bandwidth represents the upper limit of the signal bandwidth over which the channel can be considered to be flat. Beyond this the fading is considered frequency selective.

2.2.3.6 Composite descriptions of fading

Whilst undertaking research into fading for this report, the author has repeatedly found it difficult to effectively internalise the relevant criteria into a coherent description of the types of fading exhibited by a particular channel. Indeed, some workers seem to convey the impression that there is only a single type of fading present for each type of scenario. In an effort to consolidate understanding, a summary of the criteria are presented below prior to display in a graphical form after Rappaport [84, Chapter 5] in the hopes that it may similarly assist the reader.

Small scale fading due to multipath delay spread. For flat fading,

1. *Signalbandwidth* $B_S <$ *Channel bandwidth* B_C
2. *Delay spread* $\sigma_\tau <$ *Symbol period* T_S

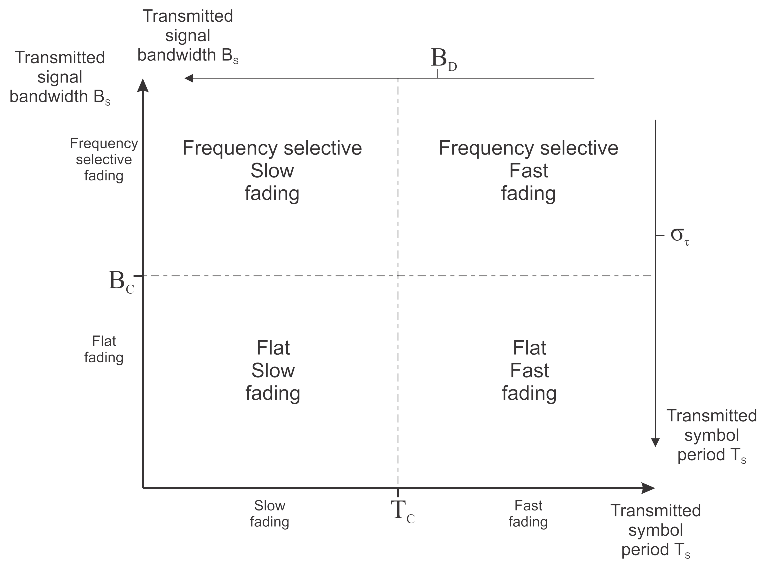

Figure 2.2: Graphical representation of small-scale fading criteria

Otherwise, the fading is frequency selective.

Small scale fading due to Doppler spread. For fast fading,

1. *Coherence time* $T_C <$ *Symbol period* T_S
2. *Doppler spread* $B_D <$ *Signal bandwidth* B_S

2.2.4 Describing the wireless channel

The previous sections have shown that there are numerous contributory factors which impair the ability of a data channel to transfer data. In this section standard expressions for the data transmission rate through a channel are reviewed, after [91, Section 3.4]. These measures will be utilised in following sections to discuss methodologies utilised to improve the data channel usage.

2.2.4.1 Nyquist Bandwidth

The theoretical upper limit for transmission over a multilevel noise-free capacity was first presented by Nyquist in the 1920s. This concept is so central to data transmission that the equation and surrounding discussions are still taught almost 100 years later:

$$C = 2B \log_2 M \tag{2.9}$$

where C is the channel capacity in bits per second, B is the channel bandwidth in Hz and M is the number of discrete signal levels available. Channel capacity is the maximum rate that data can be transmitted through a given channel.

2.2.4.2 Signal to Interference plus Noise Ratio (SINR)

Due to the material presented in the previous sections, it should now be apparent that the contributory factors to signal disruption occur as a combination of both interference and noise. This has been recognised by many wireless telecommunications workers and appears superior to Signal to Noise Ratio (SNR). The traditional measurement of signal quality appears to have been superceded by Signal to Interference plus Noise Ratio (SINR) and is utilised interchangeably by some authors within the literature without explanation [78]. The quantity is defined as:

$$SINR = \frac{S}{I + N} \tag{2.10}$$

where S is the desired signal power, I is the power from other sources of interference and N is the noise power. Clearly, in situations where there is no signal interference, the above expression simply reduces to that of the classic SNR.

2.2.4.3 Channel Capacity

This quantity is important since it is one of the parameters which determines the theoretical upper limit to the amount of information transferrable per unit time through a channel subject to AWGN using the Shannon-Hartley equation:

$$C = B \log_2 (1 + SINR) \tag{2.11}$$

as cited in standard networking texts [97, p72], [38, p68], [91, p119] but incorporating SINR.

2.2.4.4 Spectral Efficiency

Spectral Efficiency is a measure of how efficiently the channel spectrum is being utilised. The maximum achievable value is often cited thus:

$$\text{Spectral Efficiency} = \frac{C}{B} \tag{2.12}$$

2.2.4.5 Signal Energy per Bit to Noise Power Density Ratio, E_b/N_0 and Bit Error Rate (BER)

This quantity gives an effective measure of how strong a signal is. It translates to the SNR for a digital system. Whilst SNR is dependent upon the bandwidth, E_b/N_0 is not. Furthermore, E_b/N_0 can be related to a more common measure of communication quality, the Bit Error Rate (BER) for a variety of signal modulation formats (see for example, [91, Chapter 5], [100, Chapter 6], [78, Chapter 11]). The relationships are non-trivial since the analytical expressions contain a reference to the complementary error function and so conversion between the two quantities is most commonly undertaken by reference to graphical conversion charts.

2.2.5 Link budget

The previous section has outlined the mechanisms for disruption of a wireless signal, as well as the quantities to parameterise the quality of the link in general and the data-carrying capacity in particular. In order to design a practical system one of the principal calculations to be undertaken beforehand is that of link budget. Link budget calculations are used to ensure that there is sufficient power provided to a transmitter to ensure that sufficient signal power can be recovered at the receiver with an acceptable data transfer rate. For ease of calculation, quantities are usually considered logarithmically. To illustrate the basic principles consider the quantities in the simplistic expression below for a direct line of sight far-field connection with a clear first Fresnel zone:

$$P_{\mathrm{RX}} = P_{\mathrm{TX}} + G_{\mathrm{TX}} - L_{\mathrm{TX}} - FSPL + G_{\mathrm{RX}} - L_{\mathrm{RX}} \tag{2.13}$$

where:

- $\mathrm{P_{RX}}$ = Receiver power
- $\mathrm{P_{TX}}$ = Transmitter power
- $\mathrm{G_{RX}}$ = Transmitter gain (from antenna and amplifier)
- $\mathrm{L_{RX}}$ = Transmitter losses (from connections and circuitry)
- FSPL = Free Space Path Loss (2.2)
- $\mathrm{G_{RX}}$ = Receiver gain (from antenna and amplifier)

- L_{RX} = Receiver losses (from connections and circuitry)

To establish a link, the Receiver Power must be greater than the receiver detection threshold and sufficient to provide an acceptable channel capacity based on the quantities outlined in section 2.2.4.3. The reliability of the link is described by the fading margin, which for the link budget is effectively the difference (in dBm) between Receiver Power and Receiver Sensitivity and reflects the excess capacity of the link to respond to increases in disruption, most likely due to fading. Note also that a more "sophisticated" model would need to take account of fading losses and any effects arising from violated assumptions (such as the presence of obstacles within the first Fresel zone or multipath effects described by the fading margin).

2.2.6 Summary

By reference to the Shannon-Hartley expression of channel capacity (Section 2.2.4.3), increased channel capacity can be realised in situations of high bandwidth and low SINR. High bandwidth results in reduced spectral efficiency. To reduce SINR one is effectively forced to increase signal power since interference and noise contributions result from transmitting the signal itself and is not considered easy to control. Wireless signal power can be increased by the strategic deployment of carefully-chosen RF amplifiers and concentrated into a region of space using antennae. However, each antenna has a particular transmission profile which will undoubtedly overlap profiles emitted by nearby antennae to provide comprehensive wireless coverage. This overlap is likely to inhibit received signal quality due to CCI and ACI phenomena.

Multiple additional causes of signal degradation have been identified in this section as a result of the interaction of wireless signals with their environment. The greatest effects occur due to multipath and Doppler contributions to small-scale fading. Because both of these these effects each manifest to combine multiple copies of the transmitted signal together, each offset by a unique time difference to produce inter-symbol interference, at first sight the task of recovering the original signal appears Herculean at best.

Performance measures essential for the description of a communication channel have been presented and the relevance of these measures with regard to selection of appropriate signal coding techniques have been outlined.

Finally, a flavour for link budget calculation has been presented to remind the reader of the procedures that must be followed to establish a practical networked connection. As an aside, the desire for reduced wireless signal power need not be solely based upon power saving. For example, there is an increasing body of evidence upon the deliterious effects of RF radiation on living tissue (see for example [37]). In the following section measures to increase signal reliability will be reviewed, which will complete the overview of information pertinent to the physical layer.

2.3 Improving signal quality using diversity

Diversity represents the adoption of multiple signal transmissions which when processed and combined can result in a more accurate representation of the original data to be exchanged. Diversity has been highly successful in addressing much of the signal degradation issues exposed in the last section. This results in a diversity gain which represents the increase in SINR resulting from implementation of a given diversity scheme. Should different schemes be utilised for a given connection then the individual diversity gains can be combined to give a composite figure. Clearly diversity gain would provide a positive contribution to the right hand side of (2.13). This strategy of providing alternate sources of transmitted data addresses fading losses and can actually utilise multipath effects to combat fading effects: in some instances this can extend to deep fading effects of long duration. In this section an overview of the different diversity schemes are presented to illustrate the mechanisms by which substantial improvements in data rate may be accomplished. In addition, there are some alternate strategies which can also enhance signal quality in some circumstances which are also examined here.

2.3.1 Time diversity

This is effective for circumstances when the amount of fading changes with time; specifically due to Doppler fading effects from moving stations. There are two practical approaches which may be utilised. The first method is to transmit multiple copies of the same signal at different delay times. This is believed to be more effective if the delay times are greater than the coherence time. The second reported method is to incorporate forward error-correcting codes within the message to be transmitted and to interleave the symbols within the message with symbols of other messages prior to transmission. In this way some portion of a damaged received message can likely be recovered with application of the included error correction. Since any fading will change with time, then the information received at different times will suffer different fades and so likely can be recovered more effectively.

Molisch [78, Section 13.2.2] refers to a third technique utilising Automatic Repeat Request (ARQ), a link layer (layer 2) technology. Upon receiving a valid data frame, the receiver notifies the transmitter so that further processing can be undertaken. If the notification is not received by the transmitter, the frame is resent. Should the receiver not enjoy the same time diversity as the receiver then this frame could be lost indefinitely, causing infinite retransmission from the transmitter. Molisch introduces caveats in the appendices [77] which suggest that the technique is little more than an idea rather than a tractible technological solution (for example specifying low noise levels for the link and providing good error protection for the feedback loop, which if implemented would likely negate the need for a diversity improvement in the first place). In the author's opinion RFC 3366 [34] adopts a more mature approach and outlines problems as well as benefits. Overall the author is of the opinion that ARQs should be implemented carefully for wireless connections since the potential for packet loss in conditions of low SINR would negate any perceived benefit from this strategy.

Time diversity addresses signal losses from burst errors, co-channel interference and Doppler fading.

2.3.2 Frequency diversity

If the spectrum being utilised is subject to frequency-selective fading, then the signal can be transmitted simultaneously on several different frequency channels which are spread over the range and separated by at least the coherence bandwidth. If the channels are too close together in frequency then inter-symbol interference may occur, effectively negating any improvement in reception and in some instances actually worsening the situation.

2.3.3 Polarization diversity

Multiple versions of a signal are transmitted and received via antennas with different polarization. The motivation here is that whilst the propagation mechanisms for the different polarisations are essentially the same, the measured small-scale fading is largely uncorrelated [58]. Furthermore, no additional bandwidth is required to upgrade the wireless link; although clearly two antennas are required since the polarised beams at the receiver must be orthogonal.

Molisch [78, Section 13.2.5] expresses the opinion that orders of polarisation diversity exceeding 2 (after Andrews et al. [22]) should be viewed with some scepticism. Intuitively, attempts to extract additional information from polarised signals should incur substantial coupling effects within the derived sources. However, the original work has been extended to result in their application and grant of a patent [23].

2.3.4 Diversity resulting from different number and placement of antennae

When engineers investigated the possibility of utilising more than a single transmit and single receive antenna for a data connection, a number of different diversity schemes and techniques were revealed. These are outlined in more detail in the following sections.

2.3.4.1 Antenna diversity

Different antenna arrangements can significantly affect the performance and throughput of wireless links. Essentially, the signal is transmitted over several different propagation paths in an attempt to mitigate the effects of deep fading by utilising multipath effects. Antenna diversity is implemented by using multiple transmitter antennas (transmission diversity), multiple receiving antennas (reception diversity) or both. In the case of receiver diversity, a diversity combining technique can be applied, often requiring a compromise to be made between final signal quality and the complexity of the combination method. The literature frequently refers to these as Single Input Single Output (SISO), Single Input Multiple Output (SIMO), Multiple Input Single Output (MISO) and Multiple Input Multiple Output (MIMO). In general increasing numbers of antenna can leverage increases in performance and/or throughput whilst significantly increasing the complexity of the receiver signal processing. These are shown in Figure 2.3.

If the antennas are far apart, for example at different cellular base station sites or WLAN access points, this is called macrodiversity or site diversity. If the antennas are at a distance in the order of one wavelength, for example within a cellular handset, this is called microdiversity.

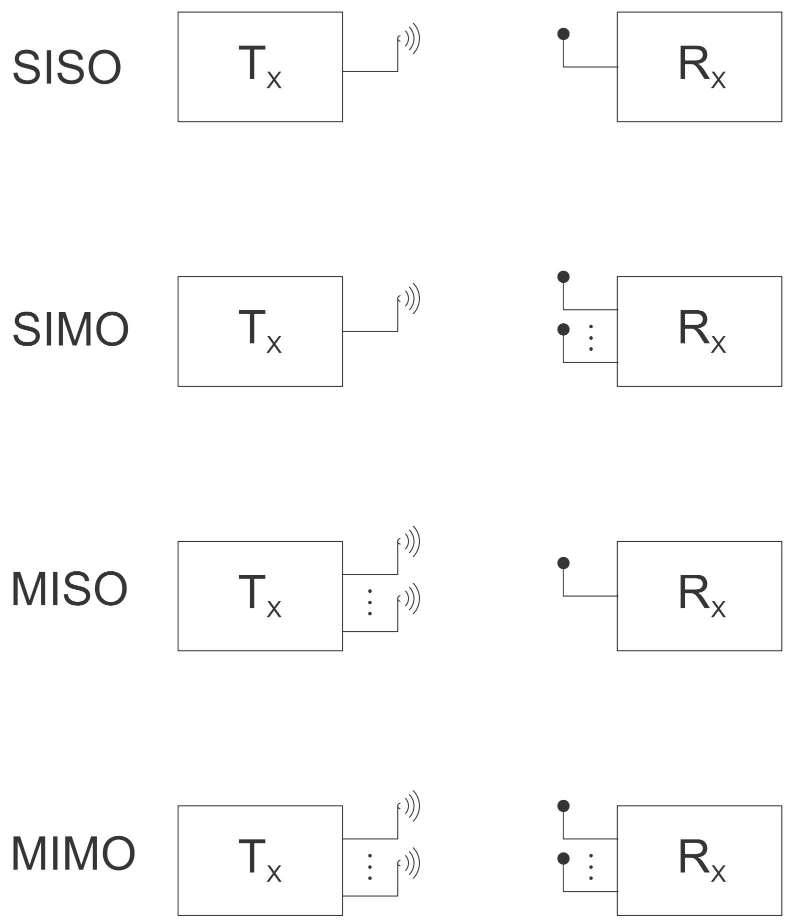

Figure 2.3: Schematic showing different antenna systems

2.3.4.2 Spatial multiplexing

Spatial multiplexing transmits different information over each separate transmission antenna to leverage an increased data transfer rate. The technique requires at least as many receiving antennas as there are transmitting antennas since the wireless data detected by a given receiver will contain a mixture of all the transmitted signals following multipath effects. In order to undertake the data transformations necessary to be able to recover the original transmitted signals, a process comparable to solving simultaneous equations needs to be undertaken, which thereby dictates the requirement.

A further modification of this technique is to introduce the concept of space-time coding. Effectively diversity is introduced into the transmission by sending differently-encoded versions of the same data stream from each transmit antenna. Reception can proceed with only a single antenna, although more are frequently deployed to improve overall performance. There are two popular decoding mechanisms. The original technique was to utilise a trellis code which fundamentally examines the most probable path through a received code-word sequence and from this deducing the transmitted symbol (see for example [78, Section 14.3]. This conveys an additional coding gain to the data stream but is complex to implement. The most popular implementation of space-time coding was first presented by Alamouti [21]. Its principal advantage over other schemes is its method for achieving transmission diversity for two antennas with no loss of available bandwidth and the ensuing simplicity in recovering the original signal which requires minimal computational effort compared to other approaches.

2.3.4.3 Beamforming

This technique utilises multiple antennas arranged in an array. By careful selection of array elements within an antenna matrix, the individually produced wireless signals can be directed to combine their output power constructively and produce high power in particular directions and very low power in others. This has the effect of concentrating the wireless energy into a focussed "beam" more effectivly than a directional antenna, thereby significantly reducing spurious energy loss in other directions.

The alternative implementation, phase-based beamforming, is highly sensitive to location. Disruption to the signal by either poor transmission of ACKs or movement of the MS for example can even produce the resulting destructive interference to be received at the MS.

Adaptive antenna arrays therefore appear more responsive and also allow for diversity gains from both polarisation diversity and spatial multiplexing. Unfortunately this technology may only be applied in some instances from the BS in a single direction in many cases, since most small-scale MS devices have support for only a single transmit antenna (see section 2.2.2) and do not always have the infrastructure to provide the implicit feedback necessary to direct the dynamic beamfinding effort.

Under conditions of true 802.11n / 802.11ac compliance, modern beamforming systems have the ability to constantly revise the signal phase to form and steer concentrated wireless signals over the best performing signal path, possibly on a per packet basis should the appropriate feedback be obtained from the receiver. This now allows multipath effects to be exploited rather than avoided. By creating a high traffic connection in this way, beamformed connections leverage antenna diversity to increase the reliability of the MS communication link. The technology is discussed in some detail in a white paper by Ruckus Wireless Inc. [87].

2.3.5 Combination of diverse signals

There are a number of methodologies available for combining received diversity signals to improve the total quality of the signal to be forwarded for data extraction. These are outlined here to illustrate the compromises between computational or electronic complexity and signal quality which may need to be considered in a wireless system.

Selection or switching diversity is by far the simplest. Here the strongest measured signal is selected from those available and passed to the receiver circuitry to be processed. All other copies are discarded.

Combining diversity frequently requires all measured signals to be demodulated before all baseband signals are collected together to form the final signal. In this case it can be difficult to produce an optimal combination. For example, one common algorithms simply adds the baseband signals together: another scales the signals to give each source an equal weight. Intuitively the method of equal weighting might be considered superior. Consider a situation where a weaker signal has a superior SINR. Here a simple addition would result in a reduced quality signal over that from the weaker high quality signal due to a combination with a source of higher SINR whereas an equal amplitude contribution would generate a better final signal. More complex combination strategies may be more or less susceptible to disruption from particular circumstances, however this is difficult to ascertain in real-time purely from the received signal strengths. Molisch [78, Section 13.4.1] examines some of these situations in more detail for the interested reader.

One particular circumstance which has not been highlighted extensively in the literature but which the author believes may be pertinent is the so-called near-far problem [44, Section 14.2.3]. To illustrate the problem: consider two transmitting stations, each transmitting with equal power with one station close to the receiver and the other far away from the receiver. Due to the dissipation by Friis' law (2.1), the received power from the closer station will be substantially greater than that from the distant source, causing the weaker signal to be masked by the stronger one and ultimately discarded within the noise level. The effects of this problem may be reduced by implementing some form of power control to ensure that received power levels are comparable for most situations.

2.3.6 Summary

The diversity techniques presented briefly above provide for higher quality information to be recovered from transmitted signals than anticipated from the additional resource expended. Whilst signal processing techniques are processor intensive and require additional power to be supplied to the device both in terms of operation and for service needs such as cooling, this is still much smaller than that which would be necessitated by the conventional approach of a higher transmission power.

When considering antenna diversity, it is important to note that the technology should be adaptive to environment. Indeed, commercial MIMO systems switch dynamically between SISO, MIMO diversity, and MIMO multiplexing modes, depending on a variety of factors including the channel environment and signal quality. For example, if the signal quality is very high the system uses spatial multiplexing, and if not, it automatically switches to spatial diversity mode or even to SISO mode. This is more important for low power considerations since MIMO intuitively utilises more power than SISO both in terms of driving the wireless signal and for the signal processing requirements at the receiver. However work by Cui and Goldsmith [31] shows that beyond some threshold value, the energy required to transmit diverse information using MIMO with Alamouti encoding is lower than the SISO alternative.

The last improvement for information recovery presented in this section was beamforming. For adaptive beamforming to occur, feedback on the signal strength must be obtained by responses from the receiving station. Whilst the process of beamforming is considered an OSI layer 1 hardware technology, its reliance on layer 2 and feedback from higher layers provides an inconvenient disruption to a clean layer-based interpretation of the technology.

2.4 Shared wireless communication

In providing a wireless infrastructure, it is only to be anticipated that providers would want to provide as many users as possible with as high quality a service as possible. In the first instance, the available spectrum can be broken down into a series of wireless channels. However, a given wireless channel will offer excess capacity at least some of the time. Since a wireless channel comprises a physical bus network, mechanisms must be found and implemented to share the available spectrum to accommodate more users and allocate ever more bandwidth to them. The accepted procedure is to adopt a MAC protocol. MAC protocols define rules for orderly access to the shared medium, providing fairness, efficiency and a methodical approach for avoiding signal collisions at the receiver. In this section the features of the more popular MAC protocols for wireless communication will be reviewed.

2.4.1 Contention-based protocols

Contentious protocols attempt to avoid a collision but if one occurs, they implement procedures to resolve the event.

2.4.1.1 "Pure" ALOHA

This was the first application of a wireless protocol and came online for the University of Hawaii in 1970 [20] (After Tannenbaum [97, Section 4.2]). The protocol itself has only two simple rules:

1. If data is available to send then transmit.

2. If a signal is detected during transmission then a collision has occurred. Try and resend later.

The rules of this asynchronous protocol shows it to be a first attempt should it be compared to more modern variants. To illustrate, there is no notion of a fixed length message or what, “later” represents. Tannenbaum makes the assumption of a fixed length message in order to develop an analytical Poisson model to illustrate how to calculate utilisation. It was this poor utilisation of resource that saw the need for improvement within the service via its syncrhonised counterpart, slotted ALOHA, outlined below.

2.4.1.2 Slotted ALOHA

In order to improve performance the original scheme was modified to synchronise on a (sic) ‘pip’ from a central transmitter station marking the start of a new transmission opportunity, or, ‘slot’. As Tannenbaum [97, Section 4.2] demonstrates, this single step more than doubled the utilisation from the previous “pure” case. However, both of these protocols highlight the need for collision detection in a contentious protocol. In the author’s opinion, the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol that followed from this work resulted in the immense success of Ethernet and ultimately leading to it becoming the single LAN standard for wired networks.

2.4.1.3 Carrier Sense Multiple Access with Collision Avoidance

Unlike Ethernet which has a restricted number of hosts on a wire, a wireless network contains a very large number of signal sources which effectively negates any adoption of collision detection functionality. Furthermore, CSMA/CD does not have the capability to address effective interaction with extended local nodes such as those described by the hidden terminal problem [98]. Whilst Tobagi and Kleinrock’s solution of a separate channel did not gain widespread adoption, in solving this with a standard OSI layer 2 Ready To Send (RTS)/Clear To Send (CTS) exchange on the single communication channel [26] creates a further problem: the exposed terminal problem. Both of these preclude CSMA/CD as a viable technology for wireless leading to the adoption of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), although these problems nevertheless impede communications as has been shown by Jayasuriya et al. [51].

Instead, a wireless station with data to transmit using CSMA/CA to a central BS access point such a router or a hub, operates by first waiting for a characteristic time (the InterFrame Space (IFS)) and if no carrier activity is detected in this time, transmits a RTS with an estimate of transmission time. Once received by the access point, after a wait equal to the IFS, the BS transmits CTS, again containing the transmission time estimate. At this point all stations within range of the recipient are aware of an impending transmission and so do not attempt to transmit for that duration. The transmitter, upon receiving this message, after another wait equal to the IFS, sends the message. If the BS receives the message, following an IFS delay, an ACKnowledgement (ACK) is sent and the transmitter can be certain the data has been received upon receipt. Should this exchange not occur as indicated, the transmitter must decide whether to requeue the data but in any case, after the anticipated transmission time has elapsed, the medium becomes available for use by any station. This proactive variant of the protocol is understandably more time consuming and so not generally worthwhile for smaller data exchanges. However, the delays resulting at the contentious part of the exchange is reduced by exchanging smaller RTS, CTS and ACK messages.

If the brief reminder of the underlying steps in the sequence presented above is insufficient, Stallings gives a more detailed exposition of a simpler variant of the process using an ACK message, including a flow chart [91, Chapter 17]. One feature of note is the inclusion of a binary exponential backoff algorithm. Whilst this doubtless succeeds in its attempt to provide a reliable response under conditions of increased loading, Kaynia [54] attributes the information lost whilst awaiting access to the shared medium to its relatively poor simulated outage probability, lying somewhere between that of pure and slotted ALOHA. However, other performance measures such as those presented by Ng et al. [80] show the comparison to be more favourable to CSMA/CA. CSMA/CA also heuristically allows a prioritised messaging system via the provision of three distinct IFS times. It should be evident to the reader that shorter IFS values indicate a shorter wait time which in turn equates directly to shorter time to transmission and hence higher priority.

2.4.2 Conflict-free protocols

For these protocols, collisions are eliminated: they can never occur. More in-depth discussions regarding more specific details of these techniques are available within wireless texts (e.g. [50, Chapter 1], [44, Chapter 14]).

2.4.2.1 Time Division Multiple Access (TDMA)

As the name suggests, access to a number of user nodes is achieved by dividing time. Each user node is allocated a fixed time slice by a centralised control node. The control node will then be required to issue synchronisation pulses to ensure that the user nodes within the network coordinate their transmissions. Having established a given time slice, the user node may then insert any required information into it. From preceding disussions relating to multipath delay spread, problems may arise due to transmission delays. This may be mitigated by providing additional guard times around each time slice.

The notion of time slicing occurs for numerous technologies within Computer Science: from multiprocess, multiprocessor application deployment to long haul fibre-optic communication systems. The advantage of time slicing is its simplicity and apparent fairness (since each stakeholder in the system receives an equal time portion). However, in reality, different stakeholders require different amounts of time and have different priorities for completion. A similar issue occurs within a wireless network for nodes transmitting general data rather than say a well-defined data stream such as say voice traffic. Even from the simple description of the protocol, Time Division Multiple Access (TDMA) is not sufficiently flexible to respond to different user needs, particularly when these are likely to change with time. Therefore, higher level traffic management is required to process data communications.

2.4.2.2 Frequency Division Multiple Access (FDMA)

Frequency Division Multiple Access (FDMA) occurs when the the available channel bandwidth is divided up into multiple individual non overlapping sub-channels. Unlike TDMA no synchronisation is required since a node may access its allocated frequency channel at any time. Furthermore, nodes requiring higher bandwidth can be allocated more than one channel. However, this is effectively a static assignment in the first instance and the available bandwidth is quantised into that of a given sub-channel capacity. Also, in an effort to reduce frequency spreading and concomitant ACI, guard bands must be allocated between frequency bands.

Whilst the conceptualisation of FDMA is appealing, it is difficult to attain effective utilisation of the available resource. Unlike the highly successful Wavelength Division Multiplexing (WDM) deployment for fibre optic media, the shared wireless medium cannot be so carefully controlled. Even if communications were transmitted with a narrow frequency range, dispersive effects from the medium would cause the receiver to be unable to effectively deconde the transmission. For this reason, many early FDMA applications have now been superceded with TDMA since the higher level management alluded to in the previous section can be configured for general communication more easily.

2.4.2.3 Code Division Multiple Access (CDMA)

Unlike the first two cases which necessitate particular allocation of resource to a given node, Code Division Multiple Access (CDMA) allows time and bandwidth to be fully utilised by any given node. CDMA The receiver can then extract the messages sent by a given user using the properties of the code used to create the message.

The data stream is XORed with a high frequency stream comprising the uncorrelated code signal to produce a wide bandwidth low transmission power spread spectrum signal from the Shannon-Hartley law 2.11 and causing little CCI as a consequence.

The codeword allocated to a given node may be obtained as a column from a Walsh-Hadamard matrix (although other codes, such as those proposed separately by Gold and Kasami [83, Section 12.2-5], may be used). The form of these square symmmetric matrices were first proposed over 150 years ago. They contain elements which are exclusively $\{+1, -1\}$ and each vector is mutually orthogonal; that is to say that any two vectors obtained from such a matrix are mutually perpendicular. A Walsh-Hadamard matrix of rank $n \times n$ denoted $H_{\mathrm{n,\ n}}$ has a simple relationship between its transpose and inverse which can be summarised thus:

$$H_{\mathrm{n,\ n}} {H_{\mathrm{n,\ n}}}^T = nI_{\mathrm{n}} \tag{2.14}$$

That is to say, the transpose is n times the inverse. Therefore, a data stream comprising individual data encoded with codewords for each participating network node can be extraced linearly and in a straightforward manner by multiplying by the original coding matrix. Moreover, past work on Walsh-Hadamard matrices has led to recursive generation methodologies for larger and larger matrices (e.g. [78, Chapter 18]). This means that provision can be made for any number of nodes within the network!

> "Uh, almost. There are a few, uh, provisos, a, a couple of quid pro quos..."
> *Disney's Aladdin (1992). Genie's response to Aladdin's, "You're gonna grant me any three wishes I want, right?"*

Whilst AWGN is uncorrelated and so will be filtered by the decoding process, for increased nodes utilising the available bandwidth, the coded transmissions will also effectively contribute interference to each other, thereby degrading the overall signal quality and making it increasingly difficult to extract the messages from the received signal. For this reason, CDMA systems are referred to as being interference limited. Individual signal reception power also contributes further complications as explained previously for the near-far problem (Section 2.3.5), which can be mitigated by implementing power control practices to ensure all received signals are of comparable power.

Multipath effects cause self interference with a given signal and make data recovery more difficult since the individual signals will only be orthogonal at zero phase. Multipath artefacts to the received signal have been managed with some success by using Rake receivers, which allows a selection of multipath components to be correlated and combined together (so long as the delay is greater than the chipping time T_{C}, which represents the duration of one bit of the coding sequence).

However, this is only part of a greater problem, namely determining when to sample the input to result in a high likelihood of signal recovery. Even if the transmission signal were to be synchronised centrally, this would not remedy the situation due to the different transit times from the geographically distributed nodes. Therefore, a best effort analysis is performed by attempting to seek characteristic synchronisation sequences within the received pattern and using that to base subsequent sampling from that.

2.4.3 Summary

A selection of contention-based and conflict-free protocols have been reviewed as routes to achieving shared communication on a wireless channel. Clearly the conflict-free protocols do not carry the overhead incurred by contention and so may provide a higher overall throughput. That said, there are issues with the conflict-free protocols which need to be considered. For example, both TDMA and FDMA present fairly straightforward hardware interfaces and operate by allocating fixed resources to the local nodes. This is efficient for fixed size streaming applications such as telephony but cannot be easily adapted to say streaming video, nor is there an easy method for transferring unused capacity between nodes. Higher layer functionality must be applied to leverage this flexibility. CDMA appears more flexible from a data utilisation perspective and conveys advantages in flat fading multipath environments but is more complex to implement both in hardware and software. Multipath effects must be addressed and mitigated in order to increase the likelihood of signal recovery.

The protocols presented here are by no means exhaustive. Other popular protocols utilise 'hopping' techniques in either frequency, time or both in an effort to reduce multipath effects, in addition to novel combinations of strategies to leverage the combined benefits of a number of individual techniques. See for example, [19], [78, Chapter 17], [44, Chapter 14].

As with all aspects of wireless communication, there is no single technology that will function equally well under all circumstances. In fact, different wireless access technologies combine aspects of all the variations in leased and open wireless frequency allocation, diversity and wireless channel sharing to produce network and service architectures for differing deployments with differing degrees of commercial success. In the following section, the components of a selection of services will be reviewed to try and ascertain how and what, if any, common aspects of a wireless service are likely to carry forward in products designed to satisfy the next generation of wireless standards.

2.5 Past and present wireless services

In this section the low layer properties of a selection of wireless systems are presented in an attempt to appreciate the tangible improvements made within the industry that have been carried forward to future standardisation efforts. This is not intended to be an exhaustive survey since summary information is not always available to lend itself to comparison and whilst the author is not convinced that this is an appropriate course of action, it nevertheless appears a reasonable initial strategy at time of writing. The information accrued is presented in a tabular format with more recent standard information towards the bottom. Whilst there are multiple standards, wireless communications seem oriented towards satisfying two goals: data traffic and mobile telephony communications.

2.5.1 The cellular principle

In order to increase the coverage area for a service, a strategy must be employed which retains usability of the wireless bandwidth. The accepted solution is to employ what is now casually referred to as the cellular principle, which divides the coverage over a number of smaller cells, with a single BS providing nominal coverage for each cell area. In this way, frequencies can be reused so long as they are not causing undue Inter Channel Interference (ICI) or ACI between nearby cells. To ensure this is the case, a reuse distance may be calculated for specified parameters using link budget calculations (Section 2.2.5) along with other quantities of use to a wireless service provider such as the capacity of a given cell which can be used for overall planning of the size and distribution of cells. A series of sample calculations have been presented and discussed in terms of strategies for increasing capacity of the system infrastructure by Molisch [78, Chapter 17]. Some of these ideas will be referred to below as well as in the context of the current work programme.

One of the issues resulting from a cell-based approach needs to resolve the issue of mobile stations transferring between coverage offered by different cells. This is not a straightforward proposition since the transfer has to be seamless and maintain a specified quality of service. Just as a MS must make application to join a cell, so the cell needs to be able to refuse to either service the mobile station or to enter into more intensive data transfer operations depending upon the loading the BS is subjected to at that instant to maintain Quality of Service (QoS).

2.5.2 Wi-Fi network standards

Wi-Fi networks based on IEEE 802.11 would seem to operate with lower range than mobile networks and so the idea of the cellular principle presented above may not seem pertinent at first. However, as the convenience of wireless leads to a wider user base and more comprehensive infrastructures are provided, so the user density will increase necessitating a more formal planning and introduction of reuse policies on a local level. Currently the strategy adopted by many personnel is to locate a channel that appears to provide a respectable SINR and to configure devices accordingly.

In the past however, even the larger corporate wireless services appear to have been strongly oriented around a more rigid BSS with less overlap between coverage areas. Assuming this to be the case, summary data for different standards is presented in table 2.2.

Standard	Year	Frequency band (/GHz)	Channel bandwidth (/MHz)	Maximum data rate (/Mbps)	No. of MIMO streams	Modulation	Maximum indoor range (/m)
802.11	1997	2.4	22	2	-	DSSS, FHSS	20
802.11a	1999	5	20	54	-	OFDM	35
802.11b	1999	2.4	22	11	-	DSSS	35
802.11g	2003	2.4	20	54	-	DSSS, OFDM	38
802.11n	2009	2.4 and 5	20 40	72.2 150	4	OFDM	70
802.11ac	2013	5	20 40 80 160	96.3 200 433.3 866.7	8	OFDM	35

Table 2.2: Summary data for Wi-Fi networks. Data summarised from [82], [101]. Note that as range increases so maximum data rate will diminish as a result of noise and interference effects.

2.5.3 Mobile network standards

Global Information and Communication Technologies are effectively managed by International agreement under the auspices of the International Telecommunications Union (ITU), a special agency empowered by the United Nations (UN). In addition to managing wired and optical media, the ITU is responsible for mobile cellular access. In order to facilitate this, the ITU coordinates shared use of the radio spectrum globally. Also, this organisation works with other standards organisations such as the ISO, Internet Engineering Task Force (IETF) and Institute of Electrical and Electronic Engineers (IEEE) to set goals for future telecommunication systems.

To illustrate the kind of activities undertaken by the ITU radiotelecommunications section ITU-R, the concepts behind recommendation ITU-R M.1645 [2] was first initiated in the mid 1980s and finally matured in 1999. It was released to introduce their plans for the ITU International Mobile Telecommunications (IMT)-2000 family of products. The author found the foresight and thought contained in this relatively short document mind blowing for the first few readings! Although the current main business of mobile telecommunications is still concerned with principally voice traffic, the convergence towards data was being managed at an incredibly high level of abstraction. Strategies were being conceived for seamless handoff between wireless networks of different scales, namely between short range personal area wearable items such as wire-

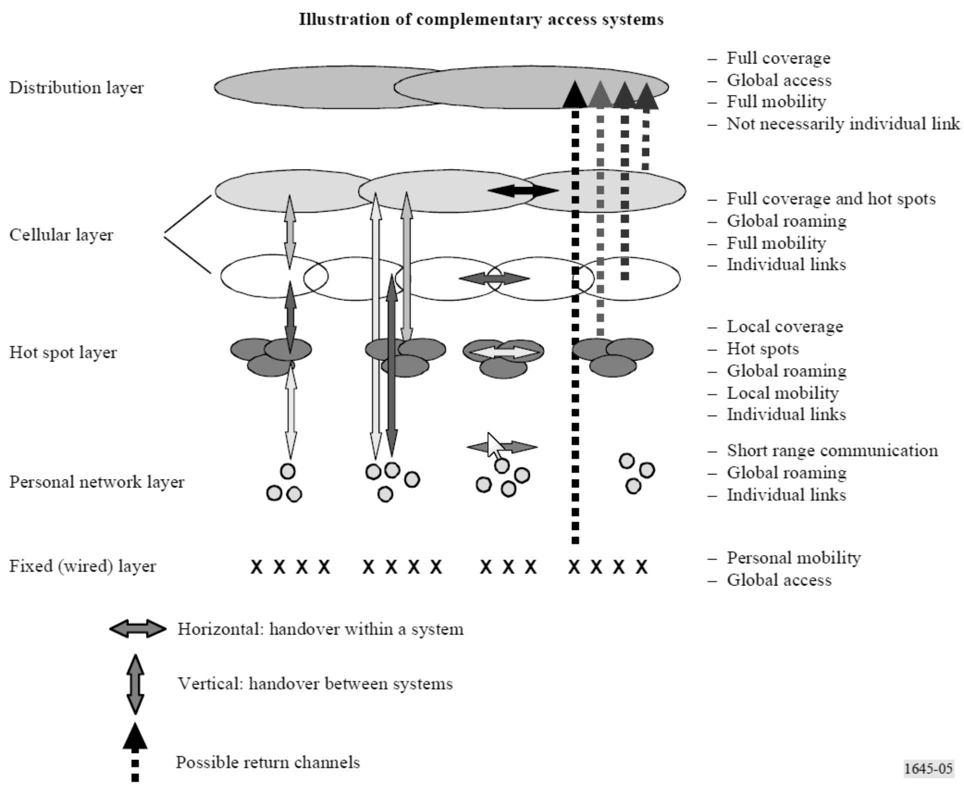

Figure 2.4: ITU IMT-Advanced Vision of Complementary Interconnected Access Systems

less earpieces within a Wireless Personal Area Network (WPAN), intermediate area items comprising video recorders and home alarm systems and wide area networks. The depth and breadth of vision extended to recommendations for planning future capacity including anticipated expansion, but also in sharing a vision of interactions between communications layers, as shown by figure 1645-05, reproduced below as figure 2.4. To try and provide an ICT reference for the state of technology when these ideas were first being conceived, the Internet as we know it was just emerging from bulletin boards and the first personal cellphones and wireless modems were appearing in stores.

In undertaking research for this work, the author has only just begun to appreciate quite what materials have been produced for wireless. When conceiving the plan for this report, the idea had been to describe the combinations of coding and modulation, diversity and multiplexing utilised in wireless telecommunications networks with the viewpoint to try and synthesise these interactions within the simulation work. The simulations would hopefully afford the opportunity to switch out block coding with a trellis code for example which would allow comparisons of both transmission performance with energy efficiency. Many of the different aspects of the work for this project have taken place in parallel, including the writing of this report. Despite having attempted to gain some appreciation for the complexities of the assorted wireless telecommunication standards the author does not feel qualified to discuss this with any confidence. In the author's opinion the materials presented in textbooks do not effectively describe the depth and richness of the strategic standardisation effort being made to bring the competing 3rd Generation standards together in compliance. This effort is probably best encapsulated within the IEEE 802.21 standard for handover between the different wireless standards [6].

The table presented in the work by Sharma [89] is most like what the author was intending to summarise the different wireless technology standards.

2.5.4 Summary

The requirement for standardisation has been shown to be far more imperative than the author could have conceived at the beginning of this project. After undertaking courses within the MINT programme the author firmly believed that the need for standards was implicit not only to provide a working methodolgy whereby implementations could produce viable reliable communication but also to gain compliance between equipment produced by different manufacturers. However, in attempting to address the standards produced by the ITU, the realisation of an inherent need to gently and continually drive disparate parties towards a consistent standardisation has exposed concepts and ideas so vast in their design that the author remains intellectually numb with awe at the vastness of the undertaking. The notion of a standardisation body bringing together numerous strong proponents, each seemingly intent on demonstrating that their technological solution is superior to all the others, towards an ultimate nirvana of interoperability and compliance is incredibly powerful: more so when one considers that if the corporations behind the different wireless solutions really are so intent on retaining their individuality they must be constantly pulling apart. It's simply incredible...

2.5.5 Discussion

As outlined in the introduction, the principal purpose of this work is to investigate strategies for reducing power consumption within wireless networks. Numerous works refer to the high energy consumption by cellphone operators which can be 20% or more of the total operating expenditure [62], [30], [73]. It is widely accepted that these figures are likely to increase as users continue to demand increased data rates to satisfy the performance requirements of more communication intensive mobile applications. To sustain market share, providers must maintain competitors' expectations and respond to the customer need to transmit an increased amount of data to an increasing number of users.

Essentially, the shared medium of a wireless network gives rise to the notion of a logical bus, insofar as all stations utilise the same free space through which electromagnetic waves travel. Wireless networks are subject to greater disruption than their wired counterparts due to the varied sources of electromagnetic interference. The most important sources of interference are usually terrestrial, lie in close geographic proximity to the station and are also most frequently associated with radio communication.

Currently all proposed performance improvements are reliant on the inferior quality of the received signal which continues to be controlled by physics and not technology. Other approaches must be sought to reduce the power consumption of wireless devices, such as those reviewed in the next section.

2.6 Energy efficiency issues

In this section a number of pertinent issues pertaining to improving the energy efficiency of ICT operations in general and wireless communication in particular are presented and discussed as a basis for selecting appropriate topics for investigation as part of this work.

2.6.1 Data centre management

Although this is not directly associated with wireless, many large institutions manage their own ICT resources in controlled environments with centralised management. Other specialised companies perform a similar activity by warehousing for smaller companies. Electronic equipment and expecially computer processors consume considerable amounts of electrical power which is dissipated as heat which needs to be extracted. Machine rooms in data centres are typically kept cool to maintain a safe operating range for the hardware by large refrigeration units. However, properly constructed modern systems are still able to perform reliably when the set temperature is increased. For example, Google recommends set temperatures of 80 °F (roughly 26.7 °C) [5]. Management services companies such as Gartner continue to raise awareness of this cost and energy saving practice in companies [76].

2.6.2 Power supply and offloading

Since the wireless handsets (MSs) are typically battery operated, their components will be selected based on their ability to operate at low power. Whilst in service, operators may not have the time to undertake a full charging cycle, nor wish to carry additional power supplies [15] and so may utilise a variety of charging mechanisms: some of which may utilise, "green" technologies [12], [14]. Consequently the equipment having the greatest opportunities for future improvement of power efficiency lies within the stationary BSs.

One straightforward improvement to BS efficiency would be to change the power source, particularly in remote areas. The best electrical conversion efficiency of a coal-fired power plant has been recorded as 49% [33]. Typically, efficiencies from raw material to electrical power socket lie at 30% which is comparable with that from a small diesel generator utilised for the majority of wireless BSs situated in remote locations. Modern engines for power production, of which the Bergen engine is probably the best in the World at time of writing, boast close to 50% efficiency [17] (although this data is representative of an engine capable of supplying power to a small city). By comparison, the best fuel cells provide power in excess of 60% efficiency at the end of system life [36].

2.6.3 Cross-layer design

Technologies to maintain reliable, high speed data transfer must utilise an approach spanning all networking layers. One topic which continues to enter the literature is the idea of a cross-layer protcol design (see for example [40], [72], [39], [44, Chapter 16]). Miao et al. [72] notes that the hardware radio interface accounts for more than half the overall system energy budget and that this proportion is likely to increase as processor sizes decrease giving lower power consumption. Further, since traditional wireless systems contain no physical layer power adaptation and consistently transmit at full power then a clear case for power regulation is evident. Other authors suggest that the existing protocol design is inflexible and sub-optimal. Some high-level functionality is sacrificed by the current protocol stacks which is particularly evident for wireless communication. One supporting argument concerns congestion control. For example, when a medium becomes congested, the standard TCP approach is that the sender does not receive ACK messages for packets sent and following some congestion control intervention, assumes that the packets have been dropped and may then decide to resend at a lower rate. Another argument concerns security. If encryption is applied at each layer, then this superfluous processing leads to an unacceptable processing overhead.

Cross-layer design can save us from all this, apparently.

Congestion control can be managed by undertaking an integrated examination at all layers. Should packet loss occur, the cross-layer design will step in and after deducing the *reason* for the dropped packets by interrogating multiple layers, modify the channel control and resend. Nowhere does any author appear willing or able to provide any metric with regard to anticipated improvement: how many packets can be salvaged by this technique under different wireless medium conditions, for example? Similarly, the authors do not report having examined the documentation of the existing utilities to see whether their proposed features can be provided by manipulation of the existing functionality.

Security can be handled by a single 'security module' that controls encryption centrally. No sensible person would simply install a single lock to the front door of their house and assume they were secure. Physical security encompasses far more than a single lock, just as digital security encompasses more than encryption. Physically, structurally sound doors and windows need to be locked when not in use and when access points are to be opened, appropriate access control needs to be employed lest the big bad wolf drops by and wants to come in... (This is of course assuming we have something a little more substantial than a house of straw) Similarly, fitting the same lock to each door and window is insufficient since once an attacker has negotiated a lock once it will be easier to break through on subsequent occasions. For this reason integrated approaches to all aspects of the entire problem are adopted and implemented in a variety of diverse forms since a sensible attacker will always attack at the weakest entry point. Security is rarely introduced at the start of a project since its complexity can exceed that of the project itself and instead is frequently left in the domain of specialists for implementation sometime after the technical aspects of project itself have been completed. For instance, the following declaration for standardising Media Independent Handover for wireless networks is not unique by any means:

> The following items are not within the scope of this standard: ... Security mechanisms [[6, Page 2]

The seminal work by Cheswick et al. [29] gives a sensible, mature and well-reasoned approach to the problem of network security that should perhaps be consulted by advocates of a centralised security module prior to submitting articles for publication.

Issues relating to layer breaking are outlined by Kawadia and Kumar [53]. These concerns do not appear to have been noted or addressed by the proponents of cross-layer design. In the author's opinion, the supporters of cross-layer design do not appear to have embraced the most basic tenets of design and have little comprehension of security. Design is a creative process so the process allows for a fluid mixing of detail and abstraction as appropriate and this follows through into software design paradigms. However, to describe a problem either a scenario is presented or more formally a use case (following [49]) showing that some thought at least has been expended in proposing how a system should respond to a situation and communicating this to others. Furthermore, one important step in any design process is to establish whether in fact the proposed product is likely to safely solve the problems it is being developed to address at some level. The authors of these works do not appear to have followed an accepted design philosophy nor have they given thought or consequence to the original reason for constructing the encapsulated software utilities they are proposing to tear down, finally showing no evidence of examining documentation to establish whether their issues can be solved within the existing framework offered by the existing code. Optimal 'clean' design procedures, although potentially considered 'clunky' under the circumstances, provide a reference frame to communicate ideas [41]. These do not appear to merit consideration for advocates of cross-layer design.

However, some authors do limit their arguments but do not seem to propose strategies for reconciling the problem. Miao et al. [72] notes improvements in power consumption achieved by specialised hardware and remarks upon substantave power saving by invoking physical shutdown intervals. The author cannot see how steps such as these would constitute layer breaking since messages could be sent down through layers whilst retaining the existing structure. The author has also been led to believe that power saving steps such as these are in fact adopted within some protocols, but this was not remarked upon. Fu et al. [40] concedes that a cross-layer design would be difficult to maintain. In the author's opinion, design elements such as well-defined interfaces and information hiding were adopted to address the very issues likely to result from implementations based on cross-layer design.

Nevertheless, some of the proposed benefits are of sufficient interest to the author that research will be undertaken on the feasibility of simulated high-level power optimisation. Should this work result in sufficient power optimisation then there may be sufficient incentive take this further. Specifically, to examine the functionality offered by different layers in more detail to see whether practical power optimising additions to the functionality could be made.

2.6.4 Summary

Whilst it is important to address the energy efficiency of wireless communication, this section has shown that strategies seemingly unrelated to wireless, for example infrastructure improvements, can produce savings in energy and expenditure. It is the author's strongest opinion that one should always examine whether improvements to a situation may be made by consideration of inputs to and outputs from the system boundary defining the problem and not solely within it. Similarly one must also critically consider the ramifications of any action at various levels of abstraction.

2.7 Summary

The environmental interactions upon transmitted wireless signals have been reviewed and have been shown to have a deliterious effect on the frequency, received power and staggered reception resulting from multipath effects. A number of standard encoding techniques used for signal transmission have been reviewed together with enhancements specific to the wireless medium. High speed low power processors leverage coding techniques which can approach the Shannon limit can retain a significant proportion of the transmitted wireless signal, which together with advances in antenna design and diversity techniques convey substantial improvements in the received data rate. Despite all this, the highest current practical commercial data rates are disappointing when compared to the original 4G standard [56], [88].

Therefore, alternate means of retaining the best possible SNR must be exploited wherever possible. International agreement with the ITU maintains particular frequency ranges for use by individual operators within each country. Each frequency range may be exclusively bought or leased, or freely shared as a service. There are in general enforceable emission limits for each use to minimise interference to other users. This have been outlined by Molisch [78, p20-21]. However, there is only a finite frequency range available over which mobile devices can operate which further confines efforts to maximising use of the available frequency allocation for a given purpose.

One strategy which has been explored within this work was conceived as a reciprocal response to Friis' law, namely that if increasing transmission power is the main control in improving signal to noise ratio and thereby data transfer rates, then an exposition of methods for increased data transfer rate may lead to a reduction in necessary transmission power. Having undertaken some background reading prior to commencing this work, the root cause for the debacle of 4th generation wireless seemed highly unsatisfactory. Specifically, the ITU-R seemed to have raised the performance standards requirements for a 4G network in March 2008 to such an extent that no wireless standard was practically able to meet it [9]. This led to an apparent reevaluation of the standard in a press release some two years later [8].

Reading around this event, articles allude to fanciful ideas of the form that the ITU bowed to pressure from manufacturers and conceding standardisation benchmarks in order to boost sales of wireless equipment. If true, then the standardisation thrust towards 5G must be some cause for concern, since there seems little technological potential for data rate improvement. From the work conducted in this study:

1. Practical coding implementations are 'close' to the Shannon limit (based on unreported and possibly incorrect calculation work conducted by the author, depending on actual channel bandwidth, practical transmitted E_b/N_0 values may be within factors between 3 and 20 of the theoretical E_b/N_0 minimum value of -1.59 dB at infinite bandwidth)

2. Current modulation techniques offer 256-Quadrature Amplitude Modula-

tion (QAM) for low noise channels. Further doubling of the data rate by a factor of 2 will necessitate an additional 256 points to be placed within the constellation diagram which must reduce the Hamming distance between states to roughly 70% of the current value. One must bear in mind that fundamentally the received signal is still subject to the same physical disturbances. These locations on the constellation diagram are dictated by signal strength. The electronic junctions may become narrower and more sensitive but they continue to respond to the same noisy signal.

3. Diversity gains appear unlikely to improve by factors in excess of 10 based on the work conducted here. Miniturisation of devices such as phased arrays and rake receivers will not convey benefit since they require distances comparable with the radiation wavelength in order to function correctly. Changing geometries by say winding receivers into a compact cylindrical shape are unlikely to improve matters significantly since the devices will continue to approach one another.

4. Multiplexing techniques also seem to be bound by the utilisation they currently receive. The only multiplexing technology which allows unlimited increases in users, namely CDMA has been suspended from future technologies in favour of 'conventional' TDMA and FDMA variants. Presumably the technical issues arising from the need for comparable received power levels from each source is the limiting cause; however capacity in the conventional multiplexing techniques cannot be extended by subdividing time and frequency channels indefinitely.

In researching the literature, the apparent technological 'leaps' leading to staggering performance improvements are not fortuitous accidents. Rather, they are the result of breakthroughs resulting from continued work on long-standing problems made possible by incremental improvements in materials processing or numerical computation on the periphery of the problems themselves. There do not appear to be any notable long-standing hurdles to be surmounted by Herculean effort remaining at the moment so in the absence of new ideas, continued performance improvements must be derived from conventional sources such as improved bandwidth at longer wavelength previously utilised for space-based communications, which is currently where the 5G data transfer rate improvements appear to originate.

The information and understanding accrued from the material presented in this section will be discussed in terms of the experimental work programme to be conducted as part of this project in the next chapter.

Chapter 3

Work programme

"I had now brought my state of life to be much easier in itself than it was at first, and much easier to my mind, as well as to my body. ... I learned to look more upon the bright side of my condition, and less upon the dark side, and to consider what I enjoyed rather than what I wanted; and this gave me sometimes such secret comforts, that I cannot express them; and which I take notice of here, to put those discontented people in mind of it..."

– Daniel Defoe, *Robinson Crusoe*

Simply from an intuitive examination of Friis' law presented in the previous section, more transmission power must be provided in order to propagate a higher power signal a longer distance. Therefore any measure which can convey a higher resultant data rate to the receiver over the same distance as an existing transmission must by definiton reduce power consumption. This improvement in reception quality could also leverage channel capacity. Consequently the initial task for this project was to attempt to ascertain what measures offer the greatest potential energy savings. Of secondary interest to the author would be work to try and establish what performance measures could be modified to provide a true 4G practical data transfer rate. The test plan likely to give consistent variation amongst candidate strategies was chosen at the onset of the work as a series of simulations using consistent data sets. The technology to be employed in these simulations was specified as MATLAB,

> "the high-level language and interactive environment used by millions of engineers and scientists worldwide"

[66]

A number of review articles have examined energy efficiency for wireless communication [28, 35, 46, 59, 94]. The most promising opportunities for improved energy efficiency indicated by these works will be discussed after the channel capacity improvement methods proposed by Molisch [78, Section 17.6.4] in section 2.5.1 have been discussed. The titles used by Molisch are presented in order below for ease of comparison. These will be commented on and discussed where appropriate in terms of providing testable metrics prior to deciding on the approach to be adopted for this part of the work.

Increasing the amount of spectrum use This is clearly a no brainer. The test would be to try and decide whether the financial outlay for the additional spectrum could be recouped from an end user. At present, far too many variables of a financial and psychological nature exist to determine a clear numerical basis upon which to base a test strategy.

More efficient modulation formats and coding A more efficient modulation format effectively requires a higher order modulation to convey more bits per symbol transmitted. The problem here is that higher order modulation becomes increasingly sensitive to noise and interference, thereby reducing the quality of the received signal with additional transmission overheads. Considering any improvements in terms of an isolated linear proportion increase, changing modulation format from say 256-QAM to 512-QAM results in 1 additional bit per 8 transmitted, or a 12.5% improvement in data rate. Anticipating the technical challenge to double the number of symbols that ultimately result in a moderate improvement in data transfer rate is not encouraging; not least since rolling this technology out would require equipment upgrades to BSs. To be attractive to a service operator any improvement should have an incremental lead-in. Similarly, utilising improved error correcting codes does not immediately offer gratification. Many standard texts refer to the current state of best practice coding as being close to the Shannon limit. Subject to the actual "close"ness of the coding, the return does not seem to justify the investment.

Better source coding Improvements in encoding techniques for streamed data such as audio and video have provided lossy compression of typical 20-200 times, providing relatively low-cost solutions to increased bandwidth since for improved codes. The power of the technique is that irrespective of the coding used, the decoding can readily be undertaken at speed with legacy equipment. Codecs and compression routines seem fairly mature, but there is the potential for enhancing this still further with compression based on fractal mathematics. Barnsley and Sloan [24] pioneered this work in the late 1980s and reported compression of the order 10,000 times whilst maintaining restoration times comparable with existing coding algorithms. However, their inherent greed in negotiatons for use of their work and technology effectively marginalised and ultimately eliminated their future participation in mainstream image compression applications. Whilst file compression conveys optimal performance for larger data and so would be highly beneficial for streaming audio or video, it may not be suitable for real-time compression of voice traffic - certainly fractal coding will not be explored within this work programme.

Discontinuous Voice Transmission It is the author's understanding from lectures taken during the MINT programme that multiplexing traffic from a single user speaking during a wireless conversation is already employed commercially. The author cannot visualise an extension to the idea other than audio/video conferencing which must be taken as being implicitly included within the idea and so no further consideration will be made of this option here.

Multiuser detection The author is probably misinterpreting the descriptions made by Molisch and focussing far too much on the stem, 'user'. The inference made by the author is that the multiple 'users' are at disparate locations and must communicate, however the exposition of the technique appears to refer to a single BS receiver, which the author is conceptually unable to distinguish from receiver diversity. Molisch may believe that the simplifying assumptions are simplifying but to the author are quite stupifying, either in terms of realising the assumptions or the consequences of violating them. No further mention will be made of this idea since the author believes it requires in-depth research and analysis which the author does not feel is appropriate for the work being undertaken here.

Adaptive modulation and coding The idea of changing modulation and coding of the transmitted signal in order to maximise channel performance is interesting. However, the author is not quite certain how this might be accomplished practically that the receiver might become aware of any changes. Possible options might include transmitting the information on a subchannel which would sacrifice bandwidth or to transmit a description of any adaptations to be made following a delay. In the latter case, implementation becomes difficult to realise. The system must wait to determine what if any adaptations would result in a tangible improvement and possibly to ensure the proposed changes would be effective. To wait further for a prolonged time to inform the receiver increases the likelihood that the conditions have changed and that a new adaptation will prove fruitful, by which time the transmitter is committed to previous changes which have been announced to the receiver. It may be interesting to undertake some simulations to determine the boundary conditions for which this proposed method may prove worthwhile.

Reduction of cell radius This is explained as constructing additional BSs to reduce cell coverage area and described as incurring high capital cost. The author is inclined to defer to Molisch but for the fact that the work was written some years ago. Since that time, less expensive MSs have been brought to market. In addition, the ICT industry has benefited from the commercial introduction of WDM into fibre optic communications. This together with more strategic planning for data lines and widespread adoption of cable laying methods requiring less excavation [27], [86] have led to a significantly reduced final cost for utilising long haul fibre connections. Conceptually at least, producing an ESS from a number of BSS with a fibre LAN backbone may be significantly less expensive than when originally costed by Molisch!

Use of sector cells Sectorized antenna arrays comprising a number of high directional gain antennas located at the BS have been proposed by a number of authors (see for example [44, Chapter 14]) as a strategy for increasing range and reducing the overall power requirement since the transmitted energy will be directed in a more focussed beam towards the MS. The actual implementation of such a scheme is however fraught with problems from the most cursory consideration. Consider a MS moving along a path through a large sector maintained by a BS. Whereas a signal transmitted over a wide angle could be utilised with low energy efficiency, a system comprising an array of narrow beam antennas would now need to include for the anticipated location(s) of the MS. In order to do this effectively, some knowledge of location, speed and direction at a given instant need to be made available to the BS, which then has to make allowances for a change in BS trajectory. The problem is non trivial but can be solved for a high proportion of cases with substantial economies of transmitter power. This may be supplemented by extended plans for situations where the BS is moving quickly or in random directions but good reliability of the connection must be maintained. However in this case the signal would likely be suffering additional interference from Doppler effects which may degrade the connection irrespective of higher level strategies. Nevertheless, sectorization in this way give rise to additional strengths and weaknesses: for example sectorising cells allows better management of MS within a smaller cell space, but there is an increased likelihood of handoff to adjoining cells within the same physical BS leading to exchange overhead within the logical units contained within the same physical location.

Use of an overlay structure Overlaying antenna coverage conveys potential benefits by concentrating localised traffic to a smaller low range antenna with the lower user density being mangaged by a long range system operating from the same location. The low user density of the clients utilising long range antenna will also permit lower power to be used by offering restricted access to modulation methods, for example offering a small number of time slices in TDMA with other slices unavailable so requiring no transmit power. As this strategy matures, so particular antennas could have their ranges extended or reduced to undertake dynamic load balancing that adapts to changing traffic patterns. This may be more manageable at a higher level of abstraction by utilising well investigated scenarios based on particular circumstances. As the situation changes then parameters describing the scenarios could be blended to provide smooth transitions between extremes. Strategically, careful thought would be required to ensure that the flexibility offered by such a rich solution could be maintained whilst still managing the increased complexity of such a deployment.

Multiple antennas Multiple antennas leverage the benefits of directional gain and hence lower power is needed to direct the beam towards the user. However, the user may be mobile which may require additional adjacent antenna to be powered to maintain coverage. However, for stationary or slow moving MSs, reducing transmitter power to these stations would convey power savings.

Fractional loading It may be a result of incomplete understanding, but the author is unsure of the benefit that Molisch is alluding to in this point. In order to offer a small percentage of timeslots in a system with a reduced reuse distance, one can only suppose that the equipment must be clearly capable of substantially higher capacity than it is being utilised for. To provide such equipment would require a substantial capital investment that is not realising its Return on Investment (RoI) by its capabilities being dialled down, which does not make good sense for the wireless operator.

Partial frequency reuse The strategy postulated by Molisch of a central subband surrounded by conventional frequency reuse is appealing as a thought, but quickly exposes itself as a single limited application of a much more powerful generic concept. To generalise the notion of frequency reuse, one might imagine orthogonal sets of reusable frequencies being made available for deployment dynamically. For example, should some of the concepts outlined in the 'Multiple antennas' item be realised, then different frequencies would be needed to adapt to the changing cell sizes, shapes and extents. To simply pool the available frequencies and utilise them randomly according to current traffic rules is likely to lead to instabilities between cooperating cells, either leading to excess CCI, ACI or potentially resulting in a defecit of user available channels for particular cells. Having some general rules may in the first instance allow one to investigate the stressors in a variety of scenarios and subjecting the results to data mining analysis, from which more flexible rules may be constructed for removing channels from and returning channels to the proposed frequency pools, possibly extending this to Artificial Intelligence (AI) algorithms which marry with underlying traffic patterns to provide a superior QoS.

Based on the materials discussed in the review articles cited at the top of this section, some additional research prospects have been identified.

Cell breathing

Cells becoming overloaded may offload traffic to adjacent cells with overlapping coverage and in so doing, may reduce the size of their own coverage area. Cells taking up this excess traffic will subsequently need to enlarge their coverage area to serve the newly-allocated MSs. Since cells are increasing and decreasing in size, the term, 'breathing' was coined. One potential benefit would be in leveraging the transmitter / receiver hardware on adjacent BSs. This would require monitoring software to conclude that there is a high probability of sustained usage in the immediate future to justify the time and resource expended in data buffering of communications traffic whilst hand-overs are performed to the disparate underused BSs. It would be interesting to attempt to obtain some data for this and undertake modelling investigations possibly using queueing models. These could then be tested in a dynamic modelling scenario to establish the efficacy of potential cell breathing strategies.

Reduced cell sizes: micro, nano and pico cells

The literature has recently become quite animated with regard to adopting this technique as interest is growing regarding the new IEEE 802.11 ac standard with its beamforming capabilities. Borrowing directly from Friis' law and the other background material presented in the Literature Review, should the transmission range halve then the power consumption can diminish by a factor of approximately 2^2 or better, depending upon the surroundings. Furthermore, intuitively a shorter transmission distance conveys less opportunity for matter to interfere with the transmitted beam, thereby reducing noise and similarly there should be fewer surfaces to produce multipath effects; all of which will act to reduce transmission power whilst maintaining high data transfer rates.

In the past, deployment of small-scale cells have been criticised due to the increased infrastructure cost. However, both wireless hardware and fibre infrastructure cost appear to be diminishing, thereby making this a potentially viable low power wireless networking solution. Furthermore, the literature has shown some work has been done with regard to effective 802.11 handover. In the past local Wi-Fi LANs have been isolated to individual departments within the corporate infrastructure or a Remote Authentication Dial In User Service (RADIUS) server environment has been deployed in larger institutions. Current advances seem to make this more attractive to smaller operations, which may lead to some interesting cultural issues with regard to access agreements in corporate real estate space.

Beamforming

Recent standards have incorporated beamforming technology which has been discussed in some detail in section 2.3.4.3. The technology does offer some performance enhancements and modelling is supported within the

MATLAB Communication System Toolbox. However the author is of the impression that combining any simulation with a ray tracing approach may be more appropriate, which may be too far removed from the general research area.

Efficient use of transmitter power

Assuming the near-far problem to have a consistent deliterious effect upon wireless reception, the faciltiy for determining transmitter power between a MS to a BS should be available. In this scenario, the BS simply utilises the minimum amount of power needed to reach a specific targeted BS rather than simply transmit using power sufficient to extend to the full coverage range.

3.1 Summary

From the suggestions presented by Molisch, a number of research opportunities have been identified and will be discussed in more specific detail below.

A simulated BS should be constructed which can maintain a number of different antennae. The antennae should have adjustable boresight, directive gain and beamwidth (Section 2.2.2) as well as transmitted power. It should be possible to overlay antennae profiles so that an appropriate MS may determine received power from all BS antennae in range. There should be the faciltity to vary the amount of coverage overlap from the antennae and to have MS monitoring for this.

The idea of cell "breathing" to dynamically accommodate the data transfer requirements will be explored in some detail. Whilst gaining an appreciation of the pertinent quantities which should be examined is clearly important, investigations should also be made into the relative sizes of cell overlap, which raises a number of issues pertaining to ACI in regions offering multiple coverage both from local overlay as well as overlaid coverage from adjacent cells, as well as developing rules with regard to simulated handover (following IEEE 802.21 [96], [4], [92]) under different fading conditons; for example, situations where a MS alternates between management by different bss due to fluctuations in monitored power need to be avoided. At time of writing, the author is considering use of a rule base similar to that encountered for a Proportional-Integral-Derivative (PID) controller local to the MS. Further logic may be exchanged by imparting a handover history to the BSs. Additional information may also be gleaned in this regard by having stations undertake random walks with changing parameters of step length and residence time following the movement.

Also, being pragmatic, a given BS may experience reduced power being supplied. The system should be able to investigate number of outages and impaired QoS to the users of a given cell.

Any simulation should be able to incorporate the effects of different coding as well as introduce simulated noise into a transmitted signal.

The system should be scalable to permit investigation of micro, nano and picocell deployments for the same space following the work undertaken by Correia et al. [30].

Representation should be made for different local environments in terms of a changing FSPL exponent as explored in section 2.1.2.

From a practical perspective, any considered option should allow an incremental upgrade procedure to be adopted, otherwise the service provider would find themselves commiting to substantial investment they would be unable to pull away from. An incremental upgrade permits changes to be made and to realise the benefits of the upgrade with a low cost extended cash flow benefit to be maintained for the service provider.

Chapter 4

Simulation studies

> *"... I could not forbear getting up to the top of a little mountain and looking out to sea, in hopes of seeing a ship; then fancy at a vast distance I spied a sail, please myself with the hopes of it, and then after looking steadily, till I was almost blind, lose it quite, and sit down and weep like a child, and thus increase my misery by my folly."*
>
> – Daniel Defoe, *Robinson Crusoe*

In this section the development procedure employed to generate the simulation data will be described together with solutions to any problems encountered. Preliminary results will be presented and discussed.

4.1 Development environment and software production

Simulation requires the generation of data upon which an analysis can be performed on measurements made from that data. As part of this work, software to enable simulation studies to be undertaken was written and utilised to establish that the proposed power reduction strategies at the BS will result in an actual power saving.

The simulation software was built to run within the MATLAB environment to allow compatability with other work being undertaken in this area. MATLAB is a weakly and dynamically typed language that is optimised to work extensively with matrices and vector data structures. Further, the MATLAB application provides a uniform interface for both console and programmatic commands. The console may be used for simple proof-of-concept investigations whilst the programmable aspect provides for repetitive data generation, simulation and analysis. When combined with extensive mathematical toolbox extensions, MATLAB provides an attractive multi-faceted scientific development environment. Unsurprisingly the environment enjoys extensive deployment in rapid prototyping scientific environments and so is a popular tool for academic research projects such as this. The software can operate with interpreted or compiled code; the student version performing 32-bit calculations in an entirely interpreted manner with its concomitant accuracy and performance issues.

The particular excitement for the author upon researching this environment was the pre-existing implementation of utilities within the Communications System Toolbox that would undertake the simulation of an entire connection, taking the original message, then performing channel coding (block or convolutional), modulation, multiplexing, etc. prior to transmission into the 'medium' using optional MIMO, at which stage noise could be introduced, with the noisy 'signal' being passed to the receiver and processed to produce a received message. This could then be compared to the original to determine metrics.

However, in order to utilise these features, first the simulation data had to be created which meant becoming familiar with MATLAB. To redress performance shortcomings throughout the product as a whole, the MATLAB environment employs a number of optimisations.

- The most notable of these is its calling convention for functions, for which MATLAB employs call by value but practically this occurs as call by reference with enforced Copy On Write (COW). COW is an optimisation method whereby data representing a passed argument is accessed from main memory via its reference (a pointer) until the argument is *written to*, at which instant a copy of the original data is made and the update applied to the new copy. The benefits of this are that processor overhead is reduced for non write argument access since virtual memory management is mitigated. However, there is no facility for releasing memory allocated to transient data upon completion of the function. In fact, there are no working utilities for memory management in MATLAB 2013 [68]. Should an application exhaust memory, the advice is to simply close and restart the application after making more memory available from the parent operating system [70]. The documentation for MATLAB memory management makes minimal recommendations and shifts responsibility for this exclusively to the developer [69].
- The programming environment provides an interactive code analyzer and optimiser. If the recommendations of this software are accepted blindly,

not only is the performance of the original code changed, but also the algorithms and from occasionally, the data it operates upon. Whilst searching for a reference to this, the author has discovered that fortunately much of this functionality has been curtailed in the most recent 2015 release. The code analyzer is no longer as agressive and also allows customisation to reporting policy for particular issues.

- The MATLAB documentation actively encourages engineers to incorporate processor and memory optimisation features within code [67–69]. In the author's opinion this can be severely detrimental to a project since each subsequent introduction of complexity to a solution also provides opportunity for errors to be made.

Following an initial attempt at producing a simulation to create a number of MS locations in direct line of sight subject only to free space loss utilising global data, some shortcomings were highlighted with regard to the distributions used. Learning from this, it was decided to programmatically insulate the development from such issues by utilising good design practices in future implementations (see for example, [41]). The simulations developed during the course of this work follow an Object Oriented (OO) paradigm in order to leverage the inherent advantages which are clearly explained. It is the author's most fervent wish to caution the reader with regard to the introduction presented in the MathWorks documentation [65] since from the author's experience this is likely to lead to unmanageable code for anything beyond a toy example and recommends a more conventional approach be adopted by the interested reader [102].

In many situations, the MATLAB environment does not yield robust solutions easily. As part of the outline research for the software development, the author learned quickly that the documentation changes from one release to another with little or no effort being invested in tracking for the end user. To further complicate matters, some aspects of the documentation do not reflect the software implementation to which it refers. Finally, there are relatively widespread reports in Internet discussion forums alluding to the poor memory management exhibited by MATLAB applications. The online documentation acts to simply return the responsibility for memory management to the developer by simply producing pages of recommendations to mitigate problems rather than avoiding them. In the author's opinion, arguably the least helpful recommendation for dealing with severe memory problems is, 'to save your work and restart MATLAB." for Windows or Linux [64, 69] operating systems.

In the author's opinion, the issues outlined above lead inexorably to code being produced which incorporates performance and memory management issues that should be handled intrinsically by any respectable enterprise strength product. This in turn results in an inferior software solution which is difficult to maintain due to the excess clutter required to manage the application environment at the expense of solving the intended engineering problems. Notwithstanding these issues together with the particular shortcomings of MATLAB from an OO perspective, the author is of the strongest opinion that a solution embracing OO principles provides a basis for mitigating the effect of many of the issues outlined above. A test-driven development paradigm [25] was adopted for development.

MATLAB memory management is based on a heap system whereby each structure requires allocation of a contiguous memory block. The system is not dynamic: there is no garbage collection and the memory does not appear to be relocatable. Allocated memory can be tagged upon allocation by reference counting, and whilst this can be used for deletion in principle, since each function is explicitly responsible for deallocation on exit, references may not be updated as the call stack unwinds and the memory allocation can be lost. This is explained in more detail by Pacula [81, Chapter 3]. Note that the "pack" function is intended to undertake some form of garbage collection but only functions from the command line and will not tidy fragmented memory allocations [63]

Software to provide a testbed for the simulations to be undertaken as part of this work was built using the OO capabilities of MATLAB within an Agile development methodology. As such, documentation of underlying ideas and concepts were embraced within the code itself rather than within separate design documents.

The memory leak problems alluded to above which were encountered whilst developing the classes for use in this work were eliminated by preallocating array storage based on the assumption that over 99% of the sample size within a population would be smaller than $4 \times \textit{standard deviation}$ greater than the mean. (For a Poisson distribution, $\textit{variance} = (\textit{standard deviation})^2 = \textit{mean}.$) The simulation was implemented by invoking a static method on a class which used a singleton instance to contain the testbed data. Following processing, the parameters obtained from the testbed were written to a file to mitigate any fragmentation from COW return variables. As the command to invoke the simulation passed out of scope, the entire object was destroyed taking allocated memory with it. This strategy was tested extensively for 1000 testbed invocations comprising in excess of 10^6 MSs with no change to the memory parameters {Maximum possible array (Limited by contiguous virtual address space available), Memory available for all arrays (Limited by virtual address space available), Memory used by MATLAB, Physical Memory (RAM)} to 4 s.f. accuracy.

A description of the algorithm encompassing the initiation of the simulation testbed using the classes created to undertake this project is given in the next section.

4.2 Algorithm description

Much of the mathematics simulating active wireless connections is based on the Poisson distribution. A Poisson distribution is a special case of a binomial distribution. Binomial distributions describe models which have two discrete states which may be selected randomly and independently (i.e. the outcome of one trial has no effect on subsequent trials) and each with a fixed probability. For a variable following a Poisson distribution, the limiting case of a binomial distribution having probability p of an event that is (close to) zero and an infinite number of trials n are considered, such that the quantity np is finite. This is referred to as the Poisson parameter μ, . This mathematical description can be manipulated to generate the form of the Poisson distribution from the general Binomial case as well as to show that the Poisson parameter for such a distribution is equal to both the expectation value and also the variance.

For each of the investigations anticipated by the current work, a software representation of a space incorporating the components of a wireless network will be constructed comprising one or more BSs with configurable sector cells comprising overlaying and multiple antennae. The range of each antenna is separately configurable. This space will be populated with individually configurable MSs. Any required parameters of the configuration may then be made by calculation. Changes will then be made to the configuration and subsequent calculations performed.

The core behaviour of each investigation is to undertake a number of Monte Carlo simulations. The principal steps within each simulation will be as follows:

- The sample space is initially populated with BSs. Antennae are affixed and configured with properties comparable to a physical deployment.
- The expected number of active user MSs will be determined from the

input parameter MS user density per unit area λ and the fixed area for the simulation; e.g. as $\pi R^2 \lambda$ for a circular area, radius R.

- Within each trial, the number of active user MSs will be generated using a standard Poisson random number generator. In the general context of telephony data, the traditional analysis technique is to employ Poisson statistics, since each connection has two states (established or not established) and there are a large (ideally infinite) number of connections, each with a low probability of success which together produce a finite expectation parameter. Indeed, connections within a telephone exchange is one of the standard examples of Poisson behaviour. Therefore, for simulation purposes, the most realistic representation of users would be to randomly generate samples which when subjected to analysis, followed a Poisson distribution. This is what the Poisson random number generator was designed to do as described by a variety of authors (see for exampe [43, Chapter 5]). The parameter to generate the data samples will be determined from the expected number of active user MSs outlined above.

- Each MS will be placed at a location in space described by numbers obtained from a random number generator having a uniform distribution. For a 2-dimensional simulation, 2 random numbers would be required, whereas 3 numbers would be neccessary in a 3-dimensional model.

- Using the location of the MS, the servicing BSs and antennae can be allocated based on received signal strength, as in a physical deployment.

- Subsequent calculations are made using the MS distance relative to its BS assuming direct line of sight transmission. The quantities to be determined include total transmitted power for a BS to service all MSs, calculated using Friis' law. SNR and Spectral Efficiency (SE) can be calculated from an estimation of received power and a supplied value of noise power. The model can provide an insight to performance within a space containing obstacles by changing the exponent for the FSPL. Finally, for the initial simulations the model would undertake a simple simulation of a power limited situation to each BS. The BS is allocated power as a proportion of that required for full service and nominally only the closest MSs would be serviced until all available power has been exhausted. This would provide data on the variation of number and range of the BSs serviced.

In order to produce simulation data which was intrinsically consistent when different properties were being varied, the random number seeds were managed to ensure the initial MS locations were identical for each simulation where practicable. In order to provide a baseline for situations where this could not be accomplished, simulations comprising small numbers of MSs were repeatedly computed until a sufficiently large number of MSs had been considered. This should allow the simulation data to be subject to statistical analysis as the model evolves.

Preliminary results, analysis and a brief discussion are presented in the next section.

4.3 Exploratory study

Following development and testing a series of simulations were run. In this section the results from one such simulation will be presented and subjected to analysis.

The selected data comprised a simulation performed with the following static parameters whilst varying the no. of equal beam width antenna. Total BS power values were calculated according to two different models. The simple model utilised constant power sufficient to provide minPower at the most distant location from the BS. The efficient model generated power sufficient to provide minPower at the calculated distance to the MS.

- subdivisions=1 i.e. a single regular hexagonal coverage area
- msDensity=0.001 i.e. 1 mobile station for each $1/0.001m^{-2} = 1000m^2$
- lossExponent=2.5 i.e. value of FSPL
- minPower=1e-11 i.e. threshold power receivable at BS in W
- radius=600 i.e. radius of circumscribing circle in m
- BSgain=1 i.e. gain of BS antenna
- MSgain=1 i.e. gain of MS antenna
- noisePower=1e-12 i.e. noise power for the WM in W
- lambda=0.33333 i.e. radiation wavelength in m
- nIter = 1071 i.e. No. of trials

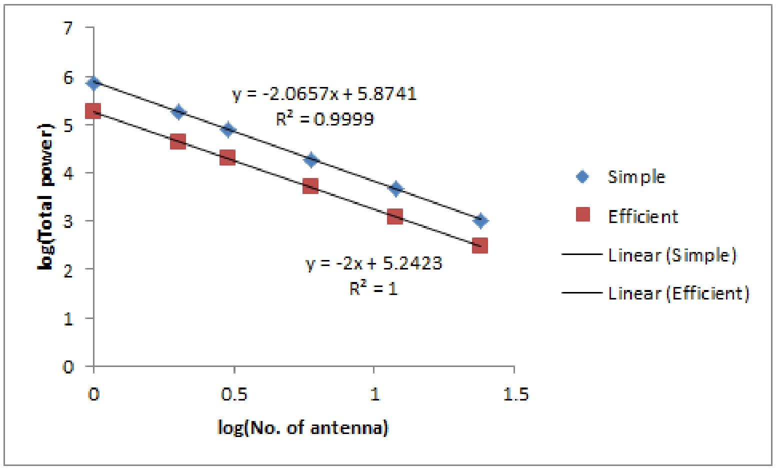

Figure 4.1: Graph showing data and fit for total BS power expended for changing number of antenna of equal half-power beamwidth

- nMS = 1001507 i.e. total mobile stations sampled

The results were transferred and tabulated into Microsoft Excel since the author finds the output easier to change than using the MATLAB figure editor. From a simple plot, the axes were changed dynamically and the properties of best fit straight lines were examined. The optimum results are shown below:

The data is explained almost exactly ($correlation\ coefficient \approx 1$) for both logarithmic ordinate ($\log_{10} y$) and abscissa ($\log_{10} x$).

Since

$$(\log_{10} y) = a + b(\log_{10} x) = \log_{10}(10^a) + \log_{10}(x^b) = \log_{10}(10^a x^b) \tag{4.1}$$

so

$$y = 10^a\ x^b \tag{4.2}$$

Both data may be summarised in the form $Power\ P = K(No.of Antenna)^n$ with parameters summarised in the table below:

			Total power		Average per BS	
	a	b	K (/W)	n	$\bar{K}$ (/W)	n
Simple	5.8741	-2.0657	748340	-2.0657	698.73	-2.0657
Efficient	5.2423	-2	174700	-2	163.12	-2

Interestingly, the exponent of the 'Efficient' transmission scheme is -2 which is indicative of an inverse square law. From the antenna section in the literature (Section 2.2.2) a causal link had been established between beamwidth and wavelength. In undertaking this test, the author had expected a value of -2.5 in keeping with the FSPL exponent which should lead to a $\lambda^{-2.5}$ and not the deduced value of -2. By comparison, the exponent obtained for the 'simple' transmission scheme is currently assumed to arise due to some unforeseen effect. No explanation for this deviation can be found within the literature. Further progress was deemed to be outside the purview of this report as advised by the project supervisor.

The code for producing the simulated data has been checked manually for simple test cases. The simulated data has been determined by alternate calculations and shown to be in agreement. Although further analyses were undertaken in light of the above interpretation the author cannot justify the presentation of potentially incorrect findings.

Chapter 5

Conclusion and future work

> *"'But,' says he again, 'if God much stronger, much might as the wicked devil, why God no kill the devil, so make him no more do wicked?' I was strangely surprised at this question; and, after all, though I was now an old man, yet I was but a young doctor, and ill qualified for a casuist or a solver of difficulties; and at first I could not tell what to say; so I pretended not to hear him, and asked him what he said; but he was too earnest for an answer to forget his question, so that he repeated it in the very same broken words as above."*
>
> – Daniel Defoe, *Robinson Crusoe*

The incentives for reducing the power consumption associated with ICT systems have been identified. Some power saving measures currently applied by a minority of institutions have been reviewed.

The physical effects of radio waves travelling through a shared medium comprising free space have been identified and interpreted in terms of causes of signal degradation. Remedial technologies implemented in wireless networking systems have been reviewed which ameliorate the effects of the environmental disturbances. A simplistic generic process of wireless signal degradation has been considered on the basis that should increased capacity reduced error communications be possible, then the potential exists for offering an acceptable lower grade service at lower power. In addition, strategies for reducing power described in current literature have been identified and where possible, critically reviewed.

The MATLAB development environment has been shown to be suitable for simulation studies related to power consumption for a wireless network. A simulation framework has been designed, constructed, tested and verified using best practice OO development paradigms. Trial experiments produced anomalous results that cannot be explained by the author and are not discussed in the literature. As a result of this, simulation experiments were discontinued until the problem is resolved.

A number of simulation experiments have been identified for testing. Some of these are in accordance with topics highlighted within review articles and so can be used for further verification once the problems relating to the simulation environment have been rectified or satisfactorily explained. Other experiments appear at the very least novel. Some experiments are oriented towards obtaining data which can be used to develop mathematical models possibly based on queueing theory or rules for switching in and changing the coverage area of BSs to actively control power. Other proposed simulations involve creating the infrastructure comparable with an actual wireless communications product and comparing this to the response when different technology implementations are replaced.

Bibliography

[1] Presidential documents. executive order 13221 - energy efficient standby power devices. *Federal Register*, 66(149):40571, August 2001.

[2] Framework and overall objectives of the future development of imt-2000 and systems beyond imt-2000. Recommendation M.1645, International Telecommunications Union, June 2003.

[3] Directive 2005/32/EC of the European Parliament and of the Council. *Official Journal of the European Union*, 191:29–58, July 2005.

[4] IEEE 802.21 tutorial, July 2006. Presented at IEEE 802.21 session 15 in San Diego, CA: http://www.ieee802.org/21/Tutorials/802%2021-IEEE-Tutorial.ppt.

[5] Google data centers, October 2008. http://www.google.ca/about/datacenters/efficiency/.

[6] IEEE standard for local and metropolitan area networks part 21: Media independent handover services, January 2009. http://standards.ieee.org/getieee802/download/802.21-2008.pdf.

[7] Cisco wireless LAN controller configuration guide, June 2010. http://www.cisco.com/c/en/us/td/docs/wireless/controller/7-0/configuration/guide/c70/c70cleanair.html.

[8] ITU paves way for next-generation 4g mobile technologies. International Telecommunications Union, October 2010. http://www.itu.int/net/pressoffice/press_releases/2010/40.aspx.

[9] Introduction to 4G. 4G Wireless Forum, 2011. http://4gwirelessforum.com/4g/.

[10] IEEE 802.11: Wireless LANs, 2012. http://standards.ieee.org/about/get/802/802.11.html.

[11] Cadence Virtuoso multi-mode simulation. Cadence Design Systems, Inc., October 2013. http://www.cadence.com/products/cic/multimode_simulation/pages/default.aspx.

[12] EZOPower universal USB solar external rechargeable backup battery, 2013. http://www.amazon.ca/EZOPower-Universal-External-Rechargeable-Battery/dp/B0076V867K/ref=cm_cr_pr_pb_t.

[13] IEEE medal for environmental and safety technologies recipients, 2013. http://www.ieee.org/about/awards/bios/envsaf$_$recipients.html.

[14] Siva cycle : Bicycle power your USB devices, 2013. http://sivacycle.com/products/atom.php.

[15] Amazon - search for "portable battery backup", October 2014. http://www.amazon.ca/s/ref=nb_sb_noss_1?url=search-alias%3Delectronics&field-keywords=portable%20battery%20backup.

[16] Clean energy ministerial: Accelerating the transition to clean energy technologies, 2014. http://CleanEnergyMinisterial.org.

[17] Integrated power solutions: Thinking outside the box. Bergen Engines: a subsidury of Rolls-Royce plc, 2014. http://bergen.rolls-royce.com/Portals/_default/assets/1/677/461_IPS2015LR.pdf.

[18] Canada's wireless policy: More choice. lower prices. better service. Industry Canada, 2015. http://www.ic.gc.ca/eic/site/ic-gc.nsf/eng/07389.html.

[19] M. Abdullah and M. F. Ghanim. An overview of cdma techniques for mobile communications. *Jounal of Mobile Communication*, 5(1):16–24, 2011.

[20] N. Abramson. The aloha system—another alternative for computer communications. In *AFIPS '70 (Fall): Proceedings of the November 17-19, 1970, Fall Joint Computer Conference*, pages 281–285, New York, NY, USA, November 1970. ACM. Library of Congress Catalog Card No.: 55-44701.

[21] S. Alamouti. A simple transmit diversity technique for wireless communications. *Selected Areas in Communications, IEEE Journal on*, 16(8):1451–1458, Oct 1998.

[22] M. Andrews, P. Mitra, and R. deCarvalho. Tripling the capacity of wireless communications using electromagnetic polarization. *Nature*, 409:316–318, 2001.

[23] M. Andrews, P. Mitra, K. Sigloch, and D. Thomson. Communicating over single- or multiple- antenna channels having both temporal and spectral fluctuations, January 2014.

[24] M. F. Barnsley and A. D. Sloan. A better way to compress images. *BYTE Magazine*, 13(1):215–223, January 1988.

[25] K. Beck. *Test Driven Development: By Example*. Addison-Wesley Professional, 1st edition, November 2002.

[26] G. Bianchi, L. Fratta, and M. Oliveri. Performance evaluation and enhancement of the csma/ca mac protocol for 802.11 wireless lans. In *Personal, Indoor and Mobile Radio Communications, 1996. PIMRC'96., Seventh IEEE International Symposium on*, volume 2, pages 392–396 vol.2, Oct 1996.

[27] N. Bradley. Installing fibre-optic cables underground. Beyond Broadband, 2012. `http://www.beyondbroadband.coop/kb/installing-fibre-optic-cables-underground`.

[28] Y. Chen, S. Zhang, S. Xu, and G. Li. Fundamental trade-offs on green wireless networks. *Communications Magazine, IEEE*, 49(6):30–37, June 2011.

[29] W. R. Cheswick, S. M. Bellovin, and A. D. Rubin. *Firewalls and Internet Security: Repelling the Wily Hacker*. Addison-Wesley Professional, 2nd edition, February 2003.

[30] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van der Perre. Challenges and enabling technologies for energy aware mobile radio networks. *IEEE Communications Magazine*, 48:66–72, November 2010.

[31] S. Cui and A. Goldsmith, Andrea.J.and Bahai. Energy-efficiency of mimo and cooperative mimo techniques in sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1089–1098, August 2004.

[32] D. Dobkin. *RF Engineering for Wireless Networks: Hardware, Antennas, and Propagation*. Elsevier Inc., 1st edition, November 2004.

[33] DONG. Avedøre power station (Avedøre værket), March 2014. `http://www.dongenergy.com/en/business%20activities/generation/activities/central_power_stations/pages/avedoere_power_station.aspx`.

[34] G. Fairhurst and L. Wood. Advice to link designers on link Automatic Repeat reQuest (ARQ), 2002.

[35] D. Feng, C. Jiang, G. Lim, J. Cimini, L.J., G. Feng, and G. Li. A survey of energy-efficient wireless communications. *Communications Surveys Tutorials, IEEE*, 15(1):167–178, First 2013.

[36] M. Fleiner. Challenges to commercialization - stationary fuel cell systems. LG Fuel Cell Systems Inc., October 2014. `http://www.keei.re.kr/keei/download/seminar/141023/II141023_b02.pdf`.

[37] M. C. for Population Health Risk Assessment. Studies examining cell phone use. University of Ottawa, 2011. `http://www.rfcom.ca/epi/cell.shtml`.

[38] B. Forouzan. *Data Communications and Networking*. McGraw-Hill Science/Engineering/Math, 5th edition, February 2012.

[39] F. Foukalas, V. Gazis, and N. Alonistioti. Cross-layer design proposals for wireless mobile networks: A survey and taxonomy. *Communications Surveys Tutorials, IEEE*, 10(1):70–85, 2008.

[40] B. Fu, Y. Xiao, H. Deng, and H. Zeng. A survey of cross-layer designs in wireless networks. *Communications Surveys Tutorials, IEEE*, 16(1):110–126, 2014.

[41] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1st edition, October 1994.

[42] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon. *Microstrip antenna design handbook*. Artech House, 1st edition, 2001.

[43] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer Science+Business Media Inc., 2nd edition, 2005.

[44] A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 1st edition, 2005.

[45] A. Grebennikov, N. O. Sokal, and M. J. E. Franco. *Switchmode RF and Microwave Power Amplifiers*. Academic Press, 2nd edition, 2012.

[46] Z. Hasan, H. Boostanimehr, and V. Bhargava. Green cellular networks: A survey, some research issues and challenges. *Communications Surveys Tutorials, IEEE*, 13(4):524–540, 2011.

[47] R. M. Islam, Y. A. Abdulrahman, and T. A. Rahman. An improved ITU-R rain attenuation prediction model over terrestrial microwave links in tropical region. *EURASIP Journal on Wireless Communications and Networking*, 2012:189:1–9, June 2012.

[48] ISO. About ISO, 2015. `http://www.iso.org/iso/home/about.htm`.

[49] I. Jacobson. *Object Oriented Software Engineering: A Use Case Driven Approach*. Addison-Wesley, July 1992.

[50] A. E. Jajszczyk. *A guide to the Wireless Engineering Body of Knowledge*. IEEE Press, 2nd edition, 2012.

[51] A. Jayasuriya, S. Perreau, A. Dadej, and S. Gordon. Hidden vs. exposed terminal problem in ad-hoc networks. In *Proc. of the Australian Telecommunication Networks and Applications Conference*, Sydney, 2004.

[52] P. Jonsson, editor. *Ericsson Mobility Report*, November 2014. `http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf`.

[53] V. Kawadia and P. Kumar. A cautionary perspective on cross layer design. *IEEE Wireless Communications Magazine*, 12(1):3–11, February 2005. ISSN 1536-1284.

[54] M. Kaynia. Performance analysis of aloha and csma in spatially distributed wireless networks. Master's thesis, Norwegian University of Science and Technology, 2007.

[55] E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley & Sons, 8th edition, 1999.

[56] L. La. 4G LTE showdown: How fast is your carrier?, August 2014. http://www.cnet.com/news/4g-lte-showdown-how-fast-is-your-carrier/.

[57] S. Lawson. LTE can mooch off of Wi-Fi spectrum with new Qualcomm chipset, February 2015. http://www.networkworld.com/article/2889793/wireless/lte-can-mooch-off-of-wifi-spectrum-with-new-qualcomm-chipset.html.

[58] W. Lee and Y. Yu. Polarization diversity system for mobile radio. *IEEE Transactions on Communications*, 20(5):912–923, October 1972.

[59] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu. Energy-efficient wireless communications: tutorial, survey, and open issues. *Wireless Communications, IEEE*, 18(6):28–35, December 2011.

[60] P. Li, N. Scalabrino, Y. Fang, E. Gregori, and I. Chlamtac. Channel interference in IEEE 802.11b systems. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 887–891, Nov 2007.

[61] J. Lifländer. Radiated efficiency: a true measure of antenna performance, November 2010. http://www.pulseelectronics.com/file.php?id=3720.

[62] D. Lister. An operator's view on green radio. In *Proceedings of the First International Workshop on Green Communications*, 2009. Keynote 3; http://www.green-communications.net/icc09/docs/GreenComm-ICC09-Keynote3-Lister.pdf.

[63] MathWorks. Command: Pack, 2014. http://www.mathworks.com/help/matlab/ref/pack.html.

[64] MathWorks. Manage workspace fragmentation as a result of long term matlab usage (32-bit windows only), 2014. http://www.mathworks.com/support/tech-notes/1100/1107.html.

[65] MathWorks. Matlab object oriented programming, 2014. http://www.mathworks.com/help/pdf_doc/matlab/matlab_oop.pdf.

[66] MathWorks. Matlab: The language of technical computing, 2014. http://www.mathworks.com/products/matlab/.

[67] MathWorks. Memory allocation, 2014. http://www.mathworks.com/help/matlab/matlab_prog/memory-allocation.html.

[68] MathWorks. Memory management functions reference, 2014. http://www.mathworks.com/help/matlab/matlab_prog/memory-management-functions.html.

[69] MathWorks. Strategies for efficient use of memory, 2014. http://www.mathworks.com/help/matlab/matlab_prog/strategies-for-efficient-use-of-memory.html.

[70] MathWorks. Resolving "out of memory" errors, 2015. http://www.mathworks.com/help/matlab/matlab_prog/resolving-out-of-memory-errors.html.

[71] A. Meier and B. Lebot. One Watt initiative: a global effort to reduce leaking electricity. In *European Council for an Energy Efficient Economy ECEEE Summer Study Proceedings, Panel 2*, 1999 Summer Study of the European Council for an Energy-Efficient Economy, Maudelieu (FR), 1999.

[72] G. Miao, N. Himayat, Y. G. Li, and A. Swami. Cross-layer optimization for energy-efficient wireless communications: a survey. *Wireless Communications and Mobile Computing*, 9(4):529–542, 2009.

[73] M. P. Mills. The cloud begins with coal - an overview of the electricity used by the global digital ecosystem. Technical report, Digital Power Group, Washington D.C., September 2013. http://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal.pdf?c761ac.

[74] S. Mingay. Green IT: A new industry shock wave. Technical report, Gartner Inc., 56 Top Gallant Road, Stamford, CT 06902. USA, December 2007.

[75] J. Miranda, R. Abrishambaf, T. Gomes, P. Goncalves, J. Cabral, A. Tavares, and J. Monteiro. Path loss exponent analysis in wireless sensor networks: Experimental evaluation. In *11th International Conference on Industrial Informatics*, pages 54–58, Bochum, Germany, July 2013. IEEE.

[76] R. L. Mitchell. IT gets its green on. *Computerworld*, 47(18):26, October 2013.

[77] A. F. Molisch. Appendix 14.a: ARQ and HARQ, 2011. `http://www.wiley.com/legacy/wileychi/molisch/supp2/appendices/c14_Appendix.pdf`.

[78] A. F. Molisch. *Wireless Communications*. John Wiley & Sons, 2nd edition, 2011.

[79] R. A. Nelson. Antennas: the interface with space, September 1999. `http://www.satellitetoday.com/publications/via-satellite-magazine/1999/09/10/antennas-the-interface-with-space/`.

[80] H.-H. Ng, W.-S. Soh, and M. Motani. On the throughput comparisons of MAC protocols in multi-hop wireless networks. *IEEE Communications Letters*, 15(12):1398–1401, December 2011.

[81] J. Pacula. Memory management in MATLAB and external C++ modules. Master's thesis, Marsarykova Univerzita Fakulta Informatiky, Czech Republic, 2012.

[82] I. Poole. IEEE 802.11 Wi-Fi standards, 2013. `http://www.radio-electronics.com/info/wireless/wi-fi/ieee-802-11-standards-tutorial.php`.

[83] J. G. Proakis and M. Salehi. *Digital Communications*. McGraw-Hill, 5th edition, 2007.

[84] T. Rappaport. *Wireless Communications Principles and Practice*. Prentice Hall, 2nd edition, 2001.

[85] S. Rice. Statistical properties of a sine wave plus random noise. *Bell System Technical Journal*, 27(1):109–157, 1948.

[86] J. Rowe. Using overhead distribution lines to carry fibre optic cables. Beyond Broadband, 2012. http://www.beyondbroadband.coop/kb/using-overhead-distribution-lines-carry-fibre-optic-cables.

[87] Ruckus Wireless. Using all the tools you can: Getting the best Wi-Fi performance possible through both beamforming and adaptive antennas, 2013. http://c541678.r78.cf2.rackcdn.com/wp/wp-using-all-the-tools-you-can.pdf.

[88] S. Segan. Fastest mobile networks 2014. PC Magazine, June 2014. http://www.pcmag.com/article2/0,2817,2459185,00.asp.

[89] P. Sharma. Evolution of mobile wireless communication networks - 1G to 5G as well as future prospective of next generation communication network. *International Journal of Computer Science and Mobile Computing*, 2(8):47–53, 2013.

[90] si Team. World to have more cell phone accounts than people by 2014. SiliconIndia Magazine, January 2013. http://www.siliconindia.com/magazine_articles/World_to_have_more_cell_phone_accounts_than_people_by_2014-DASD767476836.html.

[91] W. Stallings. *Data and Computer Communications*. Prentice Hall, 10th edition, September 2013.

[92] J. Stein. Survey of IEEE 802.21 Media Independent Handover Services, April 2006. http://www.cse.wustl.edu/~jain/cse574-06/ftp/handover/index.html.

[93] K. Stroud and D. Booth. *Advanced Engineering Mathematics*. Industrial Press Inc., 4th edition, 2003.

[94] L. Suarez, L. Nuaymi, and J.-M. Bonnin. An overview and classification of research approaches in green wireless networks. *EURASIP Journal on Wireless Communications and Networking*, 2012(1), 2012.

[95] Super-efficient Equipment and Appliance Deployment Initiative, 2014. http://www.superefficient.org/.

[96] K. Taniuchi, Y. Ohba, V. Fajardo, S. Das, M. Tauil, Y.-H. Cheng, A. Dutta, D. Baker, M. Yajnik, and D. Famolari. IEEE 802.21: Media independent handover: Features, applicability, and realization. *Communications Magazine, IEEE*, 47(1):112–120, January 2009.

[97] A. S. Tannenbaum and D. J. Wetherall. *Computer Networks*. Prentice Hall, 5th edition, September 2010.

[98] F. Tobagi and L. Kleinrock. Packet switching in radio channels: Part ii–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution. *Communications, IEEE Transactions on*, 23(12):1417–1433, Dec 1975.

[99] U.S. Department of Energy. DOE eguide for ISO 50001, 2011. `http://www.iso.org/iso/home/standards/management-standards/iso50001.htm`.

[100] M. Viswanathan. Simulation of digital communication systems using MATLAB. eBook, 2013. `http://www.gaussianwaves.com/simulation-of-digital-communication-systems-using-matlab-ebook/`.

[101] Wikipedia. IEEE 802.11, 2014. `http://en.wikipedia.org/wiki/IEEE_802.11`.

[102] Wikipedia. Object-oriented programming, April 2015. `http://en.wikipedia.org/wiki/Object-oriented_programming`.

[103] G. Wojcieszko. The role of ICT in energy efficiency. In *Proceedings of 5th ECTP-E2BA Era-net Eracobuild*, October 2011. Session 3 Keynote presentation.